BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

P410 of SEQ ID NO: 2; and an alanine to threonine mutation at a site corresponding to A485 of SEQ ID NO: 2

14. The DNA polymerase of claim 6 that has reduced discrimination against a non-conventional nucleotide selected from the group consisting of: dideoxynucleotides, ribonucleotides and conjugated nucleotides.

Please enter new claims 85-88 as follows:

85. The DNA polymerase of claim 10 wherein said mutation in Region II is selected from the group consisting of: a leucine to histidine mutation at a site corresponding to L408 of SEQ ID NO: 2; a leucine to phenylalanine mutation at a site corresponding to L408 of SEQ ID NO: 2; a proline to leucine mutation at a site corresponding to P410 of SEQ ID NO: 2.

86. The DNA polymerase of claim 10, said polymerase further comprising an alanine to threonine mutation at a site corresponding to A485 of SEQ ID NO: 2.

87. The DNA polymerase of claim 10 that has reduced discrimination against a non-conventional nucleotide selected from the group consisting of: dideoxynucleotides, ribonucleotides and conjugated nucleotides.

88. An isolated recombinant JDF-3 DNA polymerase that comprises an alanine to threonine mutation at a site corresponding to A485 of SEQ ID NO: 2.

REMARKS

As a result of this amendment, claims 1-3 and 5-47 and 85-87 are pending. Claims 4 and 11 and non-elected claims 48-84 are canceled without prejudice. Claims 1-3, 5, 46 and 47 are allowed. New claims 85-88 are added. New claims 85-87 are added to re-capture material removed from claims 12-14 by amendment to parent claim 10, and therefore add no new matter. New claim 88 is supported throughout the specification and therefore adds no new matter.

Objection to the Specification

The specification is objected to because page 23, line 5 recites "the conventional deoxynucleotides dATP, dCTP, dGTP and TTP." The Office Action states that it is believed that applicants intended that "TTP" be "dTTP," and requests appropriate correction. Applicants submit that TTP and dTTP are alternative ways of referring to the same molecule. The Examiner's attention is directed to the accompanying copy of page 957 of the 2001 Sigma-Aldrich chemical catalog (Exhibit A), which shows that thymidine-5'-triphosphate (Catalog No. T 0251) is correctly referred to as both dTTP and TTP. Applicants therefore request the withdrawal of the objection to the specification.

Rejection under 35 U.S.C. §112, first paragraph

Claims 6-15 are rejected under 35 U.S.C. §112, first paragraph for lack of written description. The Office Action states that *Thermococcus* JDF-3 is essential to the claimed invention and that the claimed organisms are not fully disclosed nor have they been shown to be publicly available. Applicants respectfully disagree.

Applicants submit that *Thermococcus* strain JDF-3 is not essential to the claimed invention. Applicants have provided polynucleotide and amino acid sequences for wild-type JDF-3 polymerase and specified sites and substitutions for a number of mutants satisfying the limitations of claims 6-15. This is all that is required to satisfy the written description requirement. Applicants submit that it is not necessary to have access to Thermococcus strain JDF-3 in order to practice the full scope of these claims, and therefore respectfully request the withdrawal of this rejection under §112, first paragraph.

Claims 6-45 are rejected under 35 U.S.C. §112, first paragraph for lack of written description on separate grounds from the deposit issue discussed above. The Office Action states that the specification only provides representative species encompassed by the claims wherein the mutant polymerase is *Thermococcus* JDF-3 and the mutation is selected from the specified residues of SEQ ID NO: 2. The Office Action states that the mutations described "are not representative of the genus of mutations claimed which encompasses any and all mutations of any Family B or *Thermococcus* species JDF-3 polymerase which results in a decrease in 3' to 5' exonuclease activity or a reduction in discrimination against non-conventional nucleotides." The Office Action also states that there is no disclosure of any particular structure to

function/activity relationship in the claimed genuses or structural characteristic other than a decrease in 3' to 5' exonuclease activity or a reduction in discrimination against non-conventional nucleotides for which no predictability of structure is apparent, concluding that the specification fails to make clear that applicants were in possession of the claimed invention at the time of filing. Applicants respectfully disagree.

The Written Description Guidelines state that the scope of the claims determines the level of detail of the written description which must be provided to support them. The specification as originally filed must convey clearly to those skilled in the art as of its filing date, that the applicant has invented the subject matter later claimed. In order to meet the written description requirement, there must be correspondence between the language of the claims and the specification. The correspondence does not have to be literal, but must be commensurate in scope. For example, if a claim covers a genus of biological materials, then the specification must describe a sufficient number of the species within the genus to convey with reasonable clarity to those skilled in the art that the inventor was in possession of the entire genus at the time the patent application was filed, or must provide guidance to permit those skilled in the art to understand what is covered by the claims.

Under the written description requirement of §112, first paragraph, the application must be reviewed in its entirety to understand what the applicant has described as the essential features of the invention (see, e.g., Wang Labs. v. Toshiba Corp., 993 F.2d 858, 865, 26 U.S.P.Q.2d 1767, 1774 (Fed. Cir. 1993). For genus claims, the written description requirement may be satisfied through a sufficient description of a representative number of species by (a) actual reduction to practice, (b) reduction to drawings, or (c) disclosure of relevant identifying characteristics, e.g., structure of other physical/chemical properties, by functional characteristics coupled with a known or disclosed correlation between function and structure, or by a combination of these sufficient to show possession of the claimed genus. The Written Description Guidelines define a "representative number of species" as sufficient species to be representative of the entire genus. A "representative number" depends on whether one skilled in the art would recognize that the applicant was in possession of the necessary common attributes or features of the elements possessed by the genus in view of the species disclosed.

Applicants submit that the specification as a whole describes the claimed invention in terms that convey to one skilled in the art that Applicants were in possession of the claimed genus at the time the patent application was filed. First, Applicants submit that, contrary to the characterization in the Office Action, the specification describes species beyond simply *Thermococcus* JDF-3 polymerase mutated at specified residues of SEQ ID NO: 2. The specification provides a listing of 55 Family B polymerases and literature references describing them (Table I), as well as accession numbers for sequence information for 17 different Family B polymerases. Applicants submit that it was known in the art that multiple sequence alignment is a means of evaluating which domains of a protein are likely to have functional significance, and that such alignment had been performed for a number of Family B DNA polymerases (see, e.g., Braithwaite and Ito, 1993, Nucl. Acids Res. 21: 787-802, and Wong et al., 1988, EMBO J. 7: 37-47, cited in the specification and incorporated therein by reference; EXHIBITS B and C). Applicants submit that alignments and knowledge of conserved regions permit one skilled in the art to identify amino acids or regions in a given Family B protein that correspond to amino acids or regions identified as critical functional determinants in *Thermococcus* JDF-3 DNA polymerase. For example, the specification teaches that Family B polymerases have six conserved structural Regions, numbered I through VI (page 7, lines 20-21), and that Region II has similar structural attributes to the nucleotide binding region of Family A polymerases, including the critical positioning of a tyrosine residue (page 7, line 21 to page 8, line 4). Thus, functional information on particular regions of Thermococcus JDF-3 Family B DNA polymerase provides a correlation between structure and anticipated function of the corresponding region in each of the Family B polymerases. The specification also provides detailed methodologies to assess the impact of a given mutation on nucleotide discrimination and 3' to 5' exo activity. Applicants describe and reduce to practice mutants of JDF-3 polymerase in the subject specification (see page 51, line 9 to page 54, line 15 and Examples 1C, 1P, 1Q and Tables V and VI), thus establishing a structure/function correlation between sites to mutate and the resulting effect of those mutations on the function of the polymerase. This structure/function correlation permits the identification of regions to mutate and types of mutations to make in other Family B polymerases in order to achieve the claimed invention using other Family B polymerases. Applicants submit that the specification thus provides a number of species that is sufficient to convey to one skilled in the art that Applicants were in possession of the full scope of the

invention at the time of filing. The methodologies for assessing function of a given Family B DNA polymerase mutant also permit the skilled artisan to determine whether a given mutant fulfills the limits of the claims.

As another example of the structure/function correlations provided by the specification, the specification also describes the relationship of Family B polymerase Region III (or Motif B) to nucleotide recognition, including description of the Region III consensus sequence structure KX_3NSXYG and its functional relationship to the $KX_3(F/Y)GX_2YG$ motif in helix O of the Family A polymerases. The specification points out that the O helix of Family A polymerases plays a role in ddNTP discrimination in Family A polymerases (page 9, line 20 to page 10, line 9). Further, the specification describes the effects of site-directed mutagenesis of Region III of Vent™ polymerase and Thermococcus barosii Family B polymerase (page 10, lines 10-23).

Applicants submit that in each instance of describing the function of various domains or residues in non-JDF-3 polymerases, the specification provides a correlation between the specific residues or regions of the non-JDF-3 Family B polymerases and the corresponding residues or regions in the JDF-3 Family B DNA polymerase. For example, in describing mutagenesis studies on Vent™ polymerase, the specification states that the studies "targeted an alanine analogous to A485 of the Thermococcus species JDF-3 DNA polymerase" (page 10, lines 10-11). As another example, this time referring to mutagenesis studies of Region II of Vent™ polymerase, the specification states "site directed mutagenesis of VENT™ DNA polymerase demonstrated that three mutations at Y412 (*which corresponds to JDF-3 DNA polymerase Y409*) could alter nucleotide binding" (page 9, lines 15-17; emphasis added). Thus, Applicants submit that the specification provides structure/function relationships for regions of Family B DNA polymerases that are reasonably applicable to all Family B polymerases.

Applicants submit that the structure/function relationships described in the specification between regions of other polymerases (both Family B and non-Family B) and regions of the JDF-3 Family B DNA polymerase, combined with knowledge in the art regarding alignment and correspondence of aligned regions, are sufficient to convey to one skilled in the art that Applicants were in possession of the claimed invention. Specifically, the specification describes 25 mutant clones of exo$^-$ JDF-3 Family B DNA polymerase (representing at least 4 different individual mutations covering both Regions II and III) and their relative nucleotide discrimination (see Tables V and VI). These mutants, together with the provided description of

structure/function relationships of Family B Regions II and III, the provided methods of testing mutants for reduced discrimination, and knowledge in the art regarding alignments and functional regions of Family B DNA polymerases provide adequate written description for the genus of isolated recombinant Family B DNA polymerases having reduced discrimination against non-conventional nucleotides wherein the polymerase has a mutation in Region II, as claimed in independent claim 10. For the same reasons, the specification provides adequate written description of the genus of recombinant Family B DNA polymerases from *Thermococcus* species JDF-3 that are 3' to 5' exonuclease deficient as claimed in independent claim 6. Also for the same reasons, the specification satisfies the written description requirement for the genus of isolated recombinant Family B DNA polymerases comprising an alanine to threonine mutation at the site corresponding to A485 of SEQ ID NO: 2 or a mutation at a site corresponding to L408, S345 or P410 of SEQ ID NO: 2, wherein the DNA polymerase has reduced discrimination against non-conventional nucleotides relative to the wild-type form of that polymerase, as claimed in independent claim 16. Because these independent claims are described in a manner meeting the written description requirement with respect to the genus claimed, it follows that the requirement is also satisfied with regard to their dependent species claims.

With regard to the scope of written description, Applicants wish to finally note that the claims reciting particular amino acid positions for mutation (e.g., claims 7-9 and 12-13, among others) recite not specific mutations of the JDF-3 Family B DNA polymerase of SEQ ID NO: 2, but mutations of amino acids "corresponding to" specific amino acids thereof. These claims therefore recognize and set forth the principle that corresponding structures or amino acids from this JDF-3 Family B DNA polymerase are applicable to other Family B DNA polymerases. Therefore, Applicants submit that the specification, of which the originally filed claims are a part, clearly describes Family B DNA polymerases in a scope broader than specific mutants of SEQ ID NO: 2, and in fact broad enough to encompass the full scope of the claims.

If the Examiner does not agree that applicants have described a sufficient number of species to fulfill the written description requirement for the genus, Applicants respectfully request that the Examiner provide Applicants with a number of species that would fulfill the written description requirement for the genus and to provide citations to case law supporting the Examiner's position. In view of the above, Applicants respectfully request the withdrawal of the written description rejection of claims 6-45.

Rejection under 35 U.S.C. §112, second paragraph

Claims 10-45 are rejected under 35 U.S.C. §112, second paragraph as being indefinite for use of the term "non-conventional nucleotides." The Office Action states that "while it is clear that dATP, dCTP, dGTP and (d)TTP are considered to be 'conventional nucleotides' it is unclear what other nucleotides, if any are also considered to be "conventional." Applicants respectfully disagree.

Applicants submit that the specification defines "non-conventional nucleotide" on page 25 of the specification as referring to "a) a nucleotide structure that is not one of *the four* conventional deoxynucleotides dATP, dCTP, dGTP and TTP recognized by and incorporated by a DNA polymerase, b) a synthetic nucleotide that is not one of *the four* conventional deoxynucleotides in (a), c) a modified conventional nucleotide, or d) a ribonucleotide (since they are not normally recognized or incorporated by DNA polymerases) and modified forms of a ribonucleotide" (emphasis added). Applicants submit that this definition makes it clear that according to the invention and with respect to DNA polymerases, there are *only* four "conventional" nucleotides, namely dATP, dCTP, dGTP and TTP. In referring to "*the four* conventional nucleotides," the definition leaves no room for more than those four conventional deoxynucleotides listed. Therefore, there is no ambiguity in the term "non-conventional nucleotides" as it (and the conventional nucleotides) is defined in the specification. Applicants respectfully request that the §112, second paragraph rejection of claims 10-45 be withdrawn.

Rejection under 35 U.S.C. §102(e)

Claims 10, 11, 14, 15 and 44 are rejected under 35 U.S.C. §102(e) as anticipated by Riedl et al., U.S. Patent No. 5,882,904. The Office Action states that Riedl et al. teaches a mutant Thermococcus barossi DNA polymerase with reduced 3' to 5' exonuclease activity and reduced discrimination against dideoxynucleotides or ribonucleotides relative to the wild type. Applicants respectfully disagree.

Applicants submit that Riedl et al. does not teach an isolated recombinant Family B DNA polymerase having reduced discrimination against non-conventional nucleotides, wherein the DNA polymerase has a mutation in Region II, as required by amended claim 10. Specifically, Riedl et al. does not teach a Family B DNA polymerase mutated in Region II. Applicants submit

that the language of the amendment "wherein the DNA polymerase has a mutation in Region II" is supported in the specification at, for example, page 52, lines 19-23. The limits of the regions of the Family B DNA polymerases, including Region II are set out in the Braithwaite and Ito and Wong et al. references (Exhibits A and B) cited in the specification and incorporated in the specification by reference. The central consensus element of Region II of the Family B DNA polymerases is also described in the specification at page 52, line 20.

Applicants submit that Riedl et al. teaches only *Thermococcus barosii* Family B DNA polymerase mutants bearing mutations in Region III, covering amino acids 488-493, and the exo mutation at amino acids 141 and 143. The reference does not teach any mutation in Region II, as required by claim 10 as amended. Therefore, applicants submit that Riedl et al. does not anticipate the invention of claim 10 and its dependents 11, 14, 15 and 44. Applicants respectfully request withdrawal of the §102(e) rejection as applied to these claims.

In view of the above, Applicants submit that all issues pertinent to patentability raised in the Office Action have been addressed herein. Applicants therefore respectfully request reconsideration of the claims.

Respectfully submitted,

Date: 1/15/02

Kathleen M. Williams
Registration No. 34,380
Palmer & Dodge, LLP
111 Huntington Avenue at Prudential Center
Boston, MA 02199-7613
Tel: (617) 239-0451
Fax: (617) 227-4420

Version of amended claims marked to show changes:

6. (Amended) An isolated recombinant Family B DNA polymerase from *Thermococcus* species JDF-3 that is 3' to 5' exonuclease deficient.

10. (Amended) An isolated recombinant Family B DNA polymerase having reduced discrimination against non-conventional nucleotides, wherein said DNA polymerase has a mutation in Region II.

12. (Amended) The DNA polymerase of claim 6 [or 10] wherein said DNA polymerase further comprises a mutation selected from the group consisting of: a leucine to histidine mutation at a site corresponding to L408 of SEQ ID NO: 2; a leucine to phenylalanine mutation at a site corresponding to L408 of SEQ ID NO: 2; a proline to leucine mutation at a site corresponding to P410 of SEQ ID NO: 2; and an alanine to threonine mutation at a site corresponding to A485 of SEQ ID NO: 2.

14. (Amended) The DNA polymerase of claim 6 [or 10] that has reduced discrimination against a non-conventional nucleotide selected from the group consisting of: dideoxynucleotides, ribonucleotides and conjugated nucleotides.

85. (New) The DNA polymerase of claim 10 wherein said mutation in Region II is selected from the group consisting of: a leucine to histidine mutation at a site corresponding to L408 of SEQ ID NO: 2; a leucine to phenylalanine mutation at a site corresponding to L408 of SEQ ID NO: 2; a proline to leucine mutation at a site corresponding to P410 of SEQ ID NO: 2.

86. (New) The DNA polymerase of claim 10, said polymerase further comprising an alanine to threonine mutation at a site corresponding to A485 of SEQ ID NO: 2.

87. (New) The DNA polymerase of claim 10 that has reduced discrimination against a non-conventional nucleotide selected from the group consisting of: dideoxynucleotides, ribonucleotides and conjugated nucleotides.

88. An isolated recombinant JDF-3 DNA polymerase that comprises an alanine to threonine mutation at a site corresponding to A485 of SEQ ID NO: 2.

EXHIBIT A

ALPHABETICAL LIST OF COMPOUNDS

PRODUCT NUMBER | US $ | PRODUCT NUMBER | US $

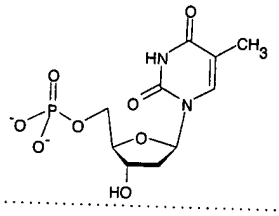
THYMIDINE 3′-MONOPHOSPHATE

T 8018 -0°C — **Ammonium Salt** — 25 mg 50.50; 100 mg 141.20
Enzymatically prepared
[40705-89-3] $C_{10}H_{15}N_2O_8P$
FW 322.2 (for free acid)

T 1008 -0°C — **Disodium Salt** — 25 mg 37.80; 100 mg 117.90
Chemically prepared
[108320-91-8]
$C_{10}H_{13}N_2O_8PNa_2$ FW 366.2

T 3512 -0°C — **Sodium Salt** — 25 mg 50.50; 100 mg 141.20
Enzymatically prepared
[108320-91-8] $C_{10}H_{15}N_2O_8P$
FW 322.2 (for free acid)

THYMIDINE 5′-MONOPHOSPHATE
(Thymidylic acid; TMP; dTMP)



T 9758 -0°C — **Free Acid** — 50 mg 14.60; 250 mg 51.70; 1 g 142.30
98-100%
[365-07-1] $C_{10}H_{15}N_2O_8P$
FW 322.2

T 7004 -0°C — **Disodium Salt** — 100 mg 12.40; 250 mg 24.20; 1 g 66.30
Minimum 99%
Also available as part of a kit.
See: Standards and Controls Section *Page 2155*
[33430-62-5] $C_{10}H_{13}N_2O_8PNa_2$ FW 366.2

THYMIDINE MONOPHOSPHATE, CYCLIC
See: Thymidine 3′:5′-Cyclic Monophosphate *Page 956*

T 4510 -0°C — **THYMIDINE 5′-MONOPHOSPHATE p-NITROPHENYL ESTER** — 25 mg 29.30; 100 mg 80.80; 250 mg 177.90
Sodium Salt
Approx. 98%
A sensitive chromogenic substrate for venom phosphodiesterase. It is not hydrolyzed by bovine spleen phosphodiesterase.
[98179-10-3] $C_{16}H_{17}N_3O_{10}PNa$ FW 465.3

T 9625 -0°C — **THYMIDINE 5′-MONOPHOSPHO-MORPHOLIDATE** — 1 g 83.20
4-Morpholine-N,N′-dicyclohexylcarboxamidine Salt
Approx. 98%
[77777-78-7] $C_{14}H_{22}N_3O_8P \cdot C_{17}H_{31}N_3O$ FW 684.8
R: 20/21/22-36/37/38 S: 26-36

THYMIDINE PHOSPHATES
See: Thymidine Mono-, Di-, or Triphosphate *Page 956*

THYMIDINE PHOSPHORYLASE
(Thymidine. orthophosphate deoxyribosyltransferase; EC 2.4.2.4)
Activity: Minimum 500 units per ml.
Unit Definition: One unit will convert 1.0 µmole each of thymidine and phosphate to thymine and 2-deoxyribose 1-phosphate per min at pH 7.4 at 25°C.
[9030-23-3]

T 2807 2-8°C — **From Escherichia coli; Recombinant Solution in 0.5 M potassium phosphate containing 2 mM uracil, 0.02% sodium azide and bovine serum albumin.** — 500 units 35.40; 1,000 units 58.90; 5,000 units 196.10

T 6632 2-8°C — **From Escherichia coli Solution in 0.5 M potassium phosphate containing 2 mM uracil, 0.02% sodium azide and bovine serum albumin. Aseptically filled** — 5,000 units 254.90

T 0251 -0°C ◆ — **THYMIDINE 5′-TRIPHOSPHATE** (dTTP; TTP) **Sodium Salt** **Approx. 97%** — 10 mg 23.40; 25 mg 43.00; 50 mg 71.30; 250 mg 278.90
[18423-43-3] $C_{10}H_{17}N_2O_{14}P_3$
FW 482.2 (for free acid)
Shipped in dry ice
R: 23/24/25-36/37/38 S: 45-26-36-22

THYMIDINE-5′-TRIPHOSPHATE-[METHYL-^{3}H]
See: Radiochemicals *Page 2148*

THYMIDYLIC ACID
See: Thymidine 5′-Monophosphate *Page 957*

T 7266 -0°C — **THYMIDYLYL(3′→5′)-2′-DEOXY-ADENOSINE Ammonium Salt** — 10 mg 83.70
[61845-39-4] $C_{20}H_{25}N_7O_{10}P$ FW 554.4 (for free acid)

T 3508 -0°C — **THYMIDYLYL(3′→5′)-2′-DEOXY-CYTIDINE Ammonium Salt** — 5 mg 51.90
[61845-38-3] $C_{19}H_{26}N_5O_{11}P \cdot NH_3$ FW 548.4

T 3883 -0°C — **THYMIDYLYL(3′→5′)-2′-DEOXY-GUANOSINE Ammonium Salt** **Approx. 95%** — 10 mg 229.50
[108321-00-2] $C_{20}H_{26}N_7O_{11}P \cdot NH_3$ FW 588.5

T 8880 -0°C — **THYMIDYLYL(3′→5′)THYMIDINE (TpT) Ammonium Salt** — 5 mg 56.20; 25 mg 196.30
[1969-54-6] $C_{20}H_{27}N_4O_{12}P \cdot NH_3$ FW 563.5

T 6633 -0°C — **THYMIDYLYL(3′→5′)THYMIDYLYL-(3′→5′)-2′-DEOXYCYTIDINE Disodium Salt** — 1 mg 41.20
[108347-86-0] $C_{29}H_{37}N_7O_{18}P_2Na_2$ FW 879.6

T 0376 RT — **THYMINE** (2,4-Dihydroxy-5-methylpyrimidine; 5-Methyluracil) **Minimum 99%** — 5 g 8.40; 10 g 12.80; 25 g 24.90; 100 g 73.50
Also available as part of a kit.
See: Standards and Controls Section *Page 2162*
See also: Tissue Culture Media and Reagents *Page 1776* and *Page 1871*
[65-71-4] $C_5H_6N_2O_2$ FW 126.1

THYMINE-2-^{14}C
See: Radiochemicals Section *Page 2148*

THYMINE-METHYL-^{3}H
See: Radiochemicals Section *Page 2148*

T 3766 2-8°C — **THYMINE 1-β-D-ARABINO-FURANOSIDE** (1-β-D-Arabinofuranosylthymine) **Minimum 99%** — 50 mg 173.00; 250 mg 576.40
[605-23-2] $C_{10}H_{14}N_2O_6$ FW 258.2

Nucleic Acids Research, 1993, Vol. 21, No. 4 787–802

Compilation, alignment, and phylogenetic relationships of DNA polymerases

Dan K.Braithwaite and Junetsu Ito*
Department of Microbiology and Immunology, College of Medicine, Arizona Health Sciences Center, Tucson, AZ 85724, USA

Received January 14, 1993; Accepted January 22, 1993

INTRODUCTION

This is an update of an earlier compilation and alignment of DNA polymerase sequences (Ito and Braithwaite, 1991). As in the previous compilation, we attempted to compile complete sequences, to facilitate the identification of conserved and viable regions of the DNA polymerases (1). This update includes, for the first time, three DNA polymerase sequences from *Archaea* (2); two new members of the Family A DNA polymerases; and 19 new members of the Family B DNA polymerases. In addition, we included nucleases that have related amino acid sequences to *E.coli* DNA polymerase I, and the sequence of *E.coli* DNA polymerase III (ε-subunit) was aligned to Family C due to its homology to *Bacillus subtilis* DNA polymerase III.

As in the previous compilation (1), Family A DNA polymerases are named for their homology to the product of the *pol*A gene specifying *E.coli* DNA polymerase I; Family B DNA polymerases are named for their homology to the product of the *pol*B gene encoding *E.coli* DNA polymerase II; and Family C DNA polymerases are named for their homology to the product of the *pol*C encoding *E.coli* DNA polymerase III alpha subunit.

Table 1 summarizes the molecular weights and isoelectric points of each DNA polymerase and nuclease. Table 1 also serves as a reference guide to the sequences shown in Figures 1A, 1B, and 1C. Since no new sequences were published for the Family X DNA polymerases (β-like), we have excluded them from this compilation.

SEQUENCE ALIGNMENT

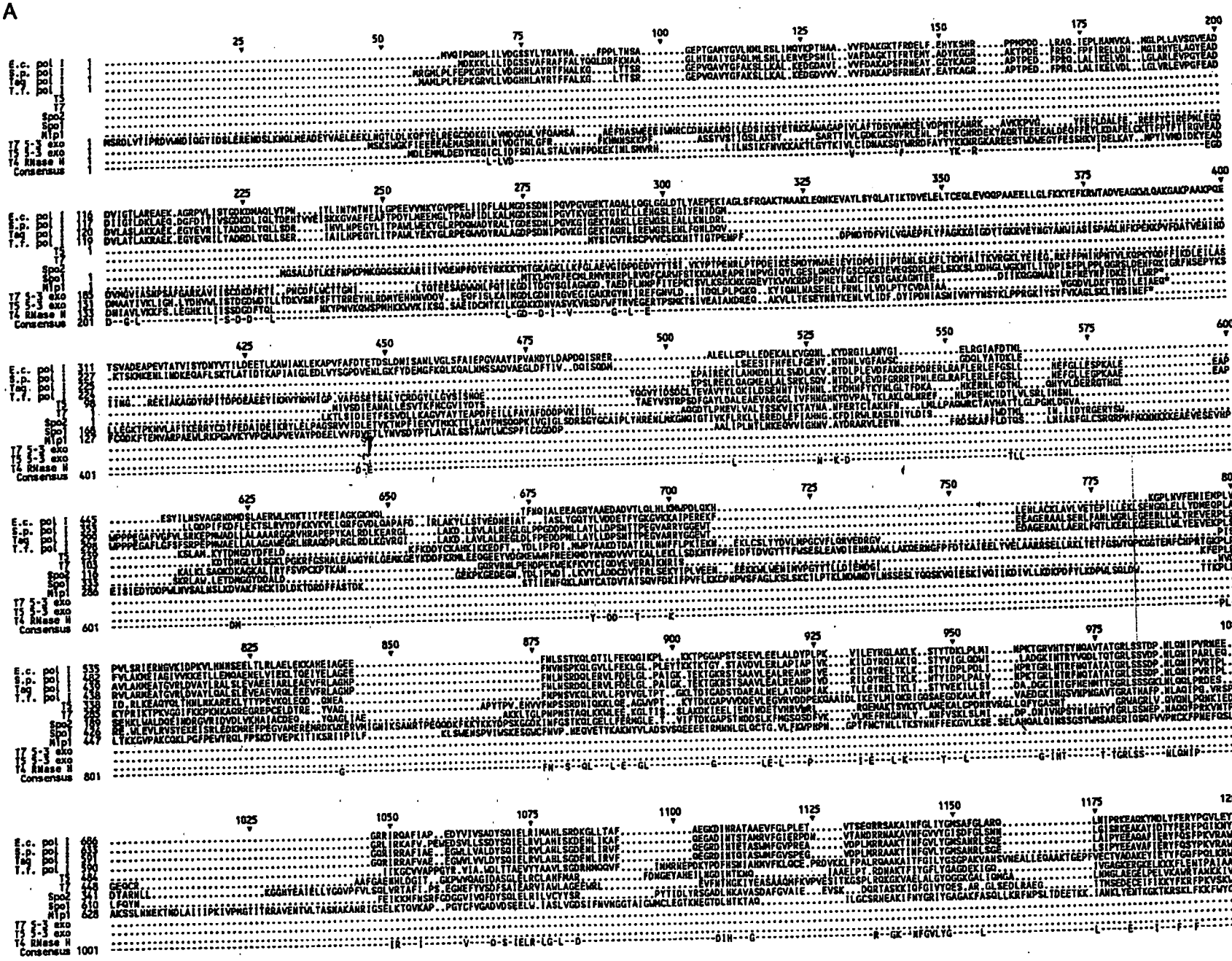
The multiple alignments of the amino acid sequences for this update were performed in most cases by merely adding on to our original alignments (1) where possible. Due to the large number of sequences added to the alignment for Family B we have changed the original alignment in some areas between obvious blocks of conserved sequences. The newer sequences were added by aligning each to the closest related sequence already aligned, or in many cases to the closest related group f sequences already aligned. A more recent addition to the UWGCG (University of Wisconsin Genetic Computer Group) program package, PILEUP, a multiple alignment program, was used extensively to try and locate significant homology in groups of closely related sequences. These newly formed groups of highly related sequences were then regapped to conform with the entire alignment based upon the previous alignment of those sequences in the new group from the original alignment. As in the previous paper, all the final adjustments had to be made by eye and, as stated above, in Family B the added sequences led to some improvements to the original alignment that became evident to the eye when they were being combined with the entire alignment by hand.

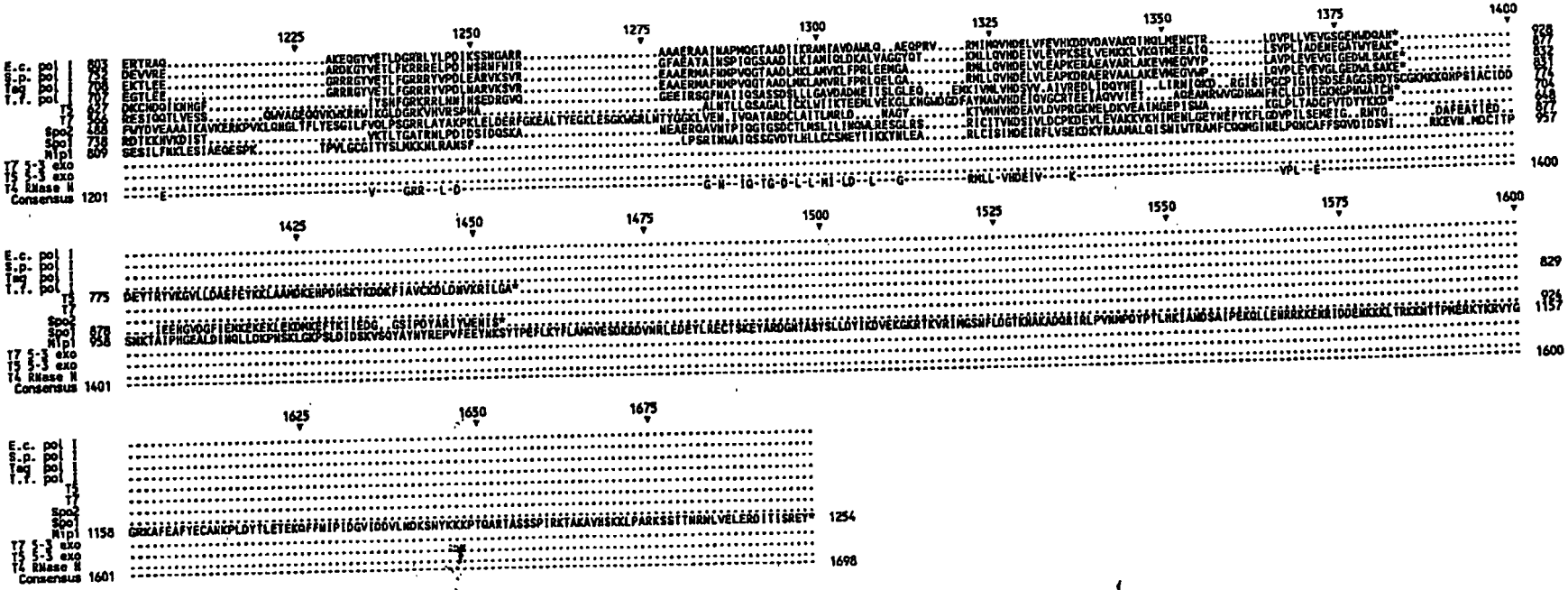
GENERATION OF PHYLOGENETIC TREES FOR THE DNA POLYMERASE DOMAINS

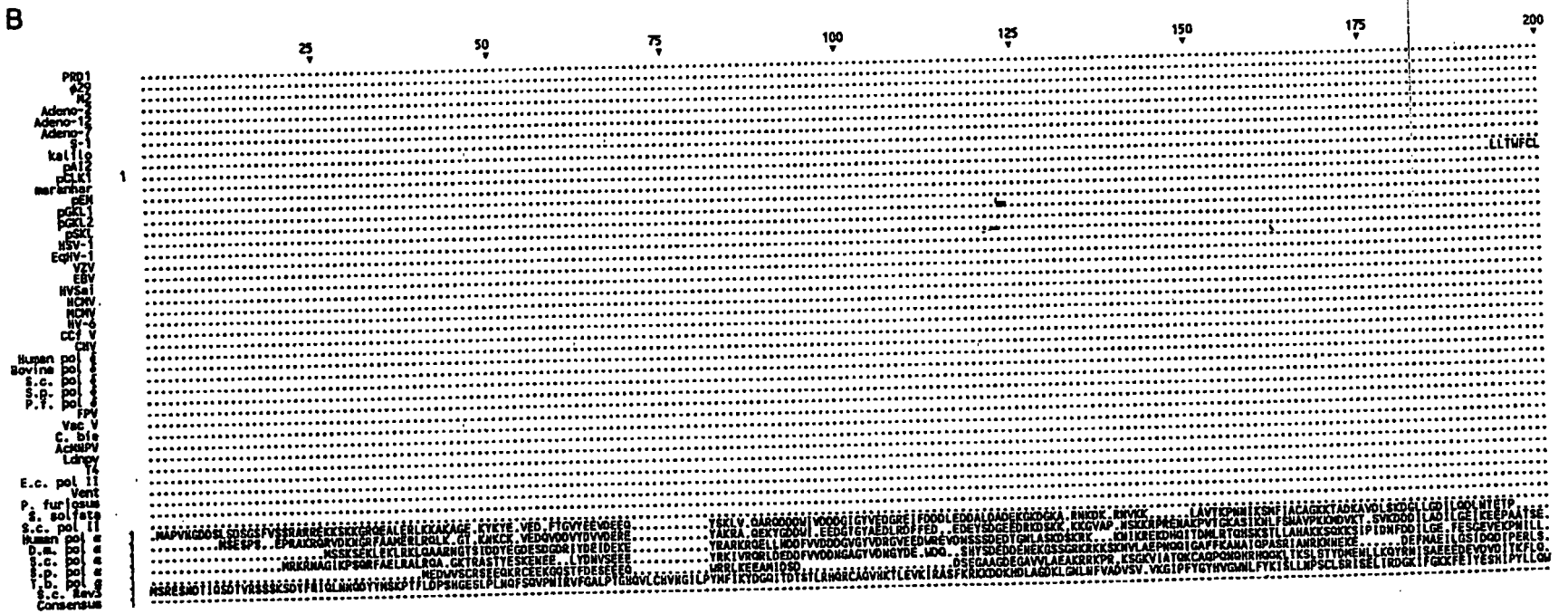
Using Felsenstein's PHYLIP program package (71), specifically the programs named in the outline below, we generated phylogenetic trees for the 9 Family A DNA polymerases (Figures 2A and 2B) and for the 47 Family B DNA polymerases (Figures 3A and 3B). The trees for Family A were created from the alignment in Figure 1A using the most conserved regions found at the following positions: 798 to 814, 877 to 998, 1047 to 1090, 1104 to 1123, 1131 to 1158, 1175 to 1206, 1236 to 1251, 1284 to 1305, 1322 to 1340, and 1365 to 1379. These conserved regions were recombined and 100 bootstrap samples were generated using SEQBOOT program. Using the DNADIST program, we turned the samples into distance matrices using the Kimura-2 parameter method. The resulting matrices were then input to the NEIGHBOR program using the UPGMA method to produce approximately 100 trees. Finally those trees were reduced to a single tree using the CONSENSE program. This final tree was then plotted for publication using two different methods. The trees in Figures 2A and 3A were created by the DRAWGRAM program setup to produce a phenogram type tree and the trees in Figures 2B and 3B were created by the DRAWTREE program. The trees for Family B were created from the alignment in Figure 1B, according to the same procedure, using the most conserved regions found at the following positions: 1407 to 1760, 1885 to 1901, 1956 to 1990, 2081 to 2100, 2181 to 2210, and 2280 to 2320. The Family B DNA polymerases can be subdivided into two subfamilies, the protein-primed DNA polymerase subfamily and the RNA-primed DNA polymerase subfamily.

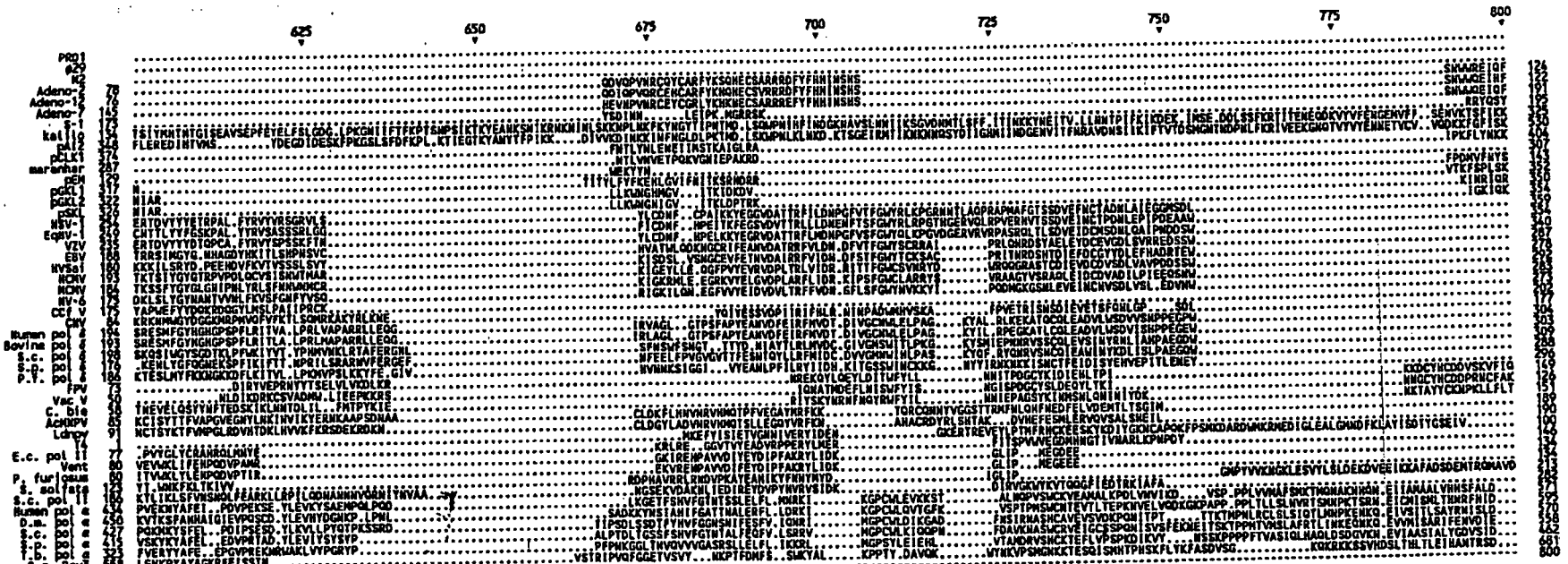
* To whom correspondence should be addressed

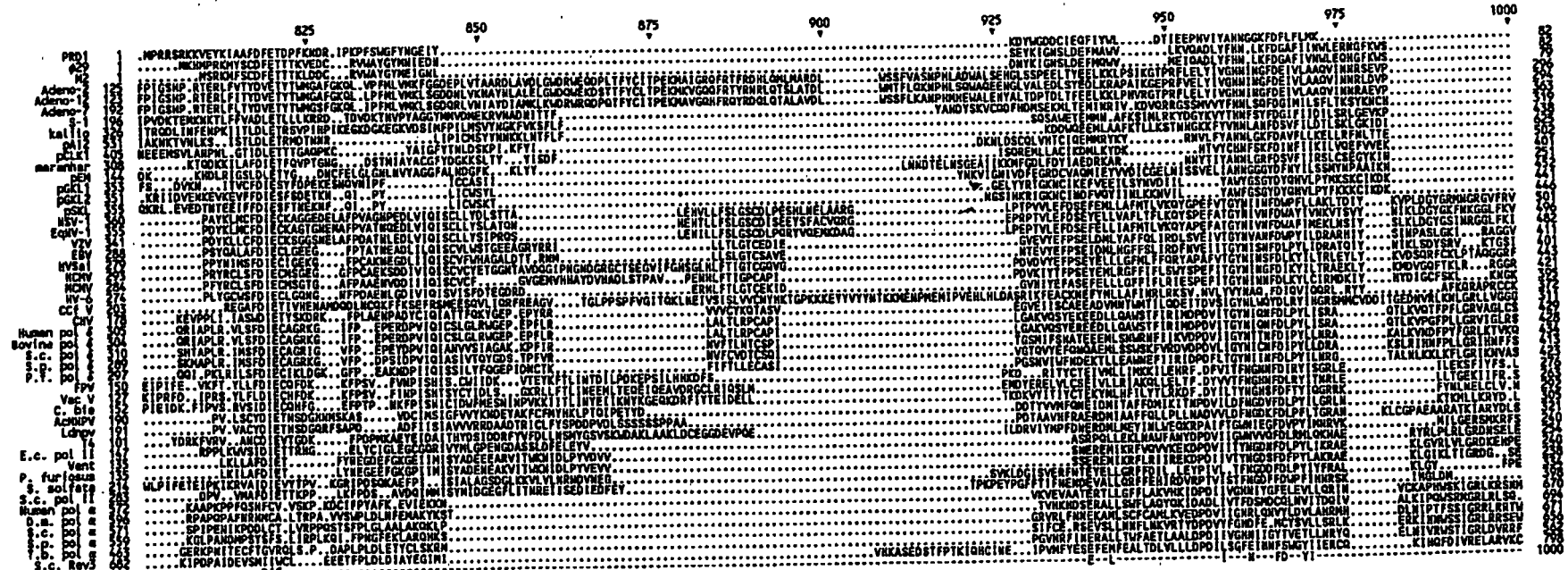
A

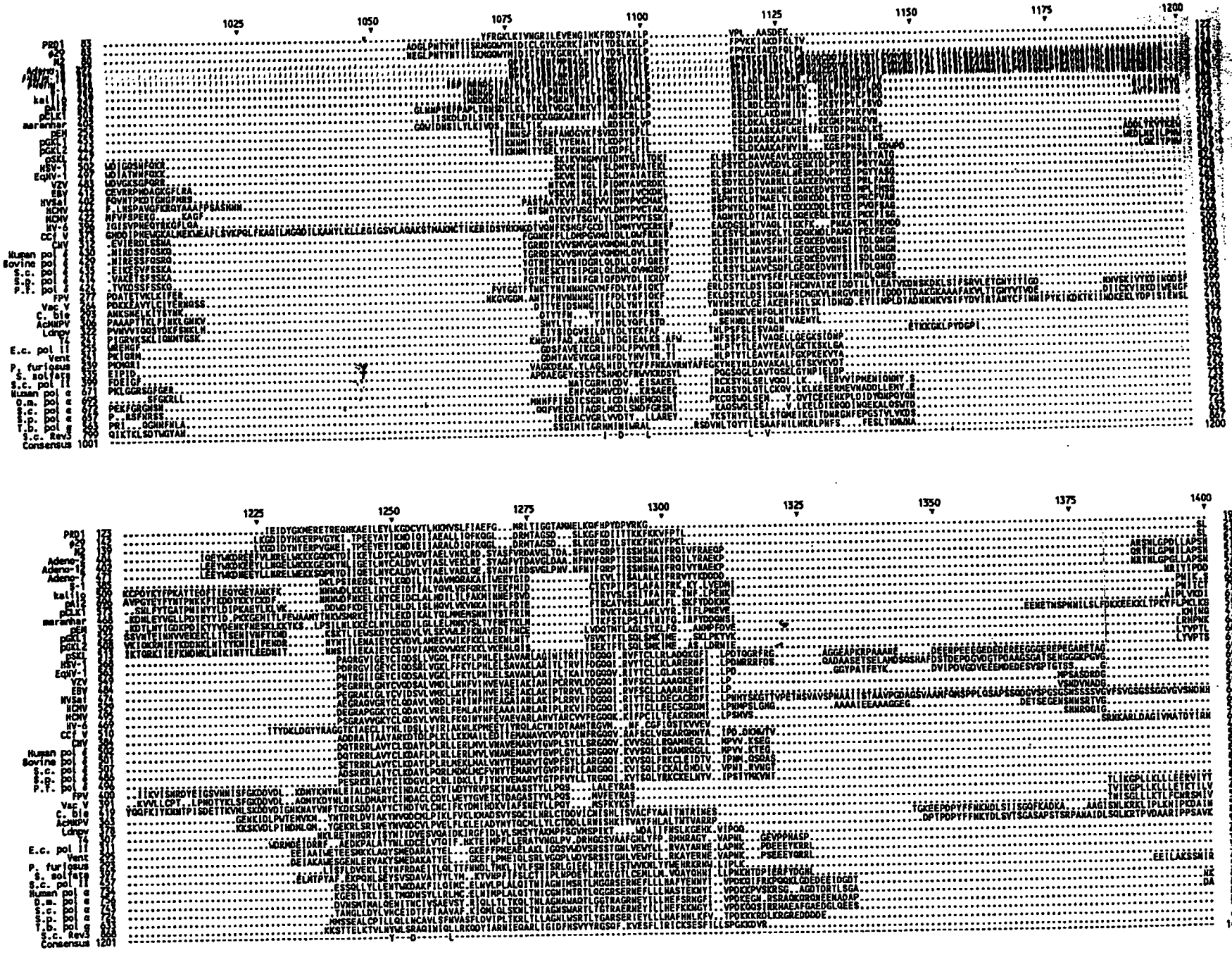


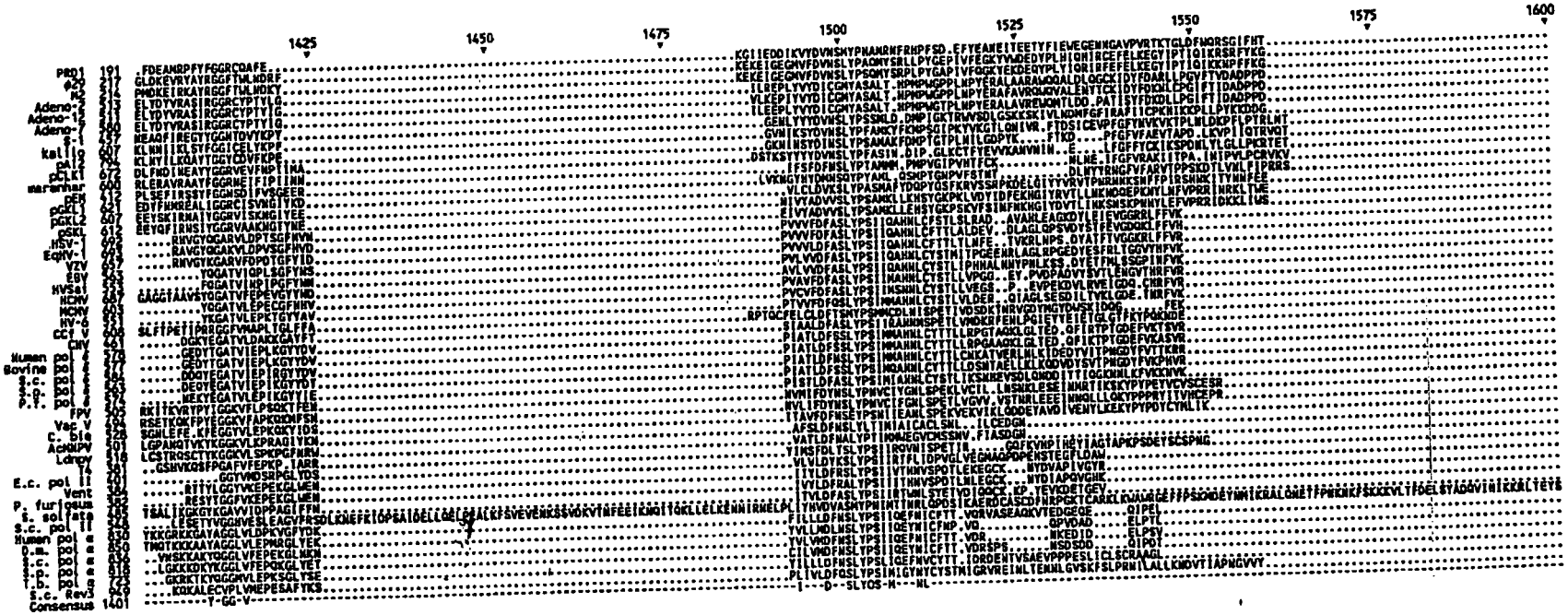


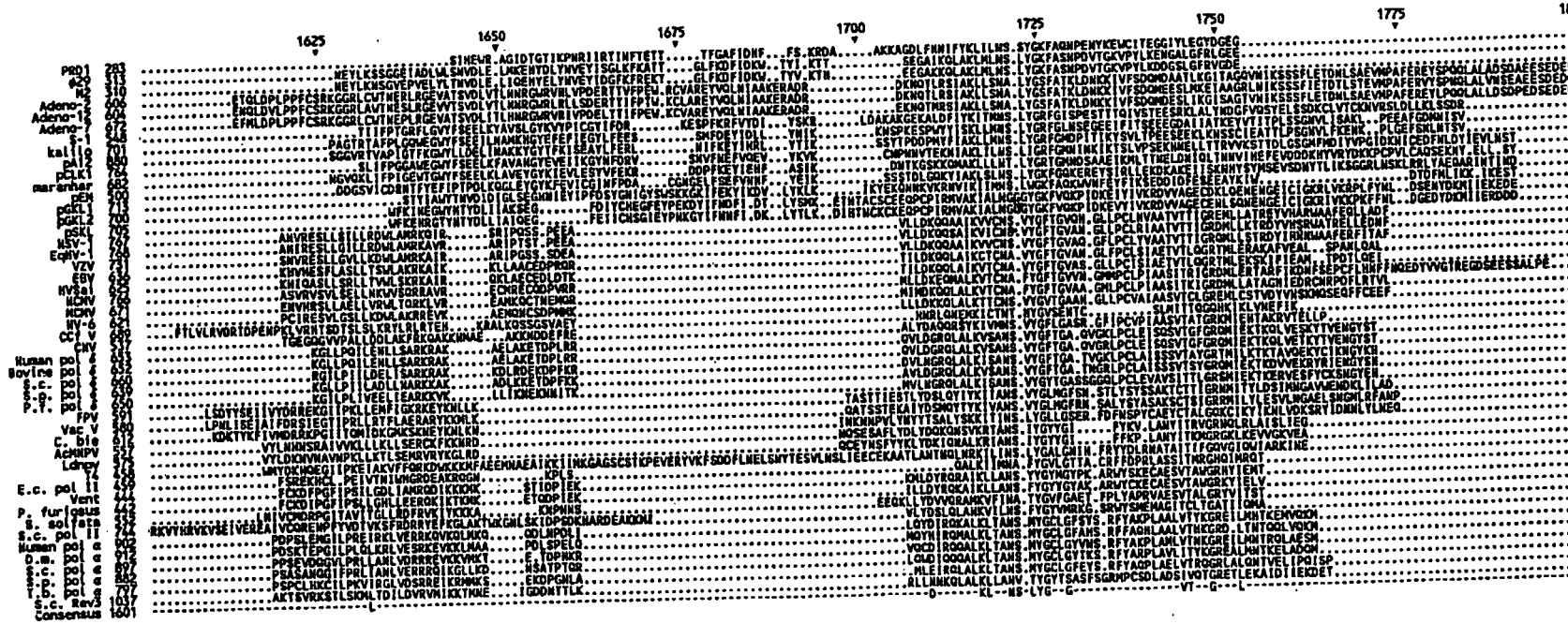


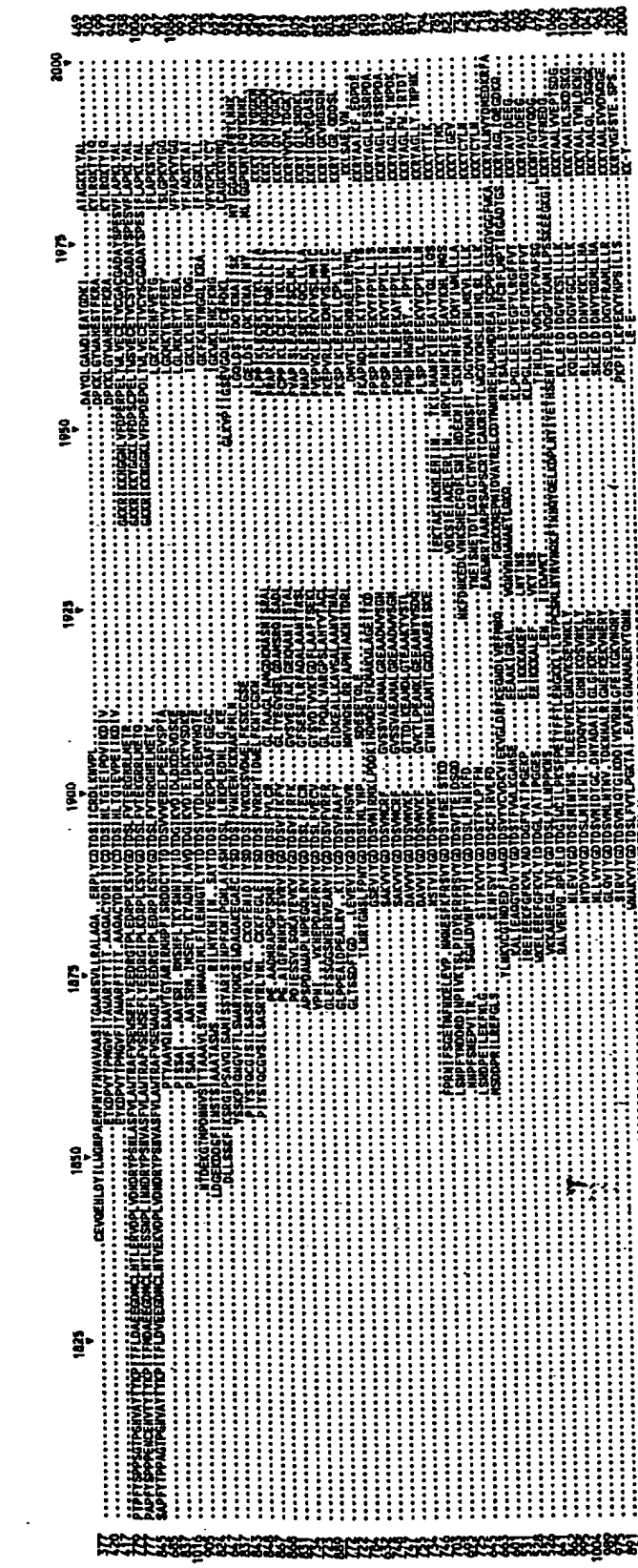


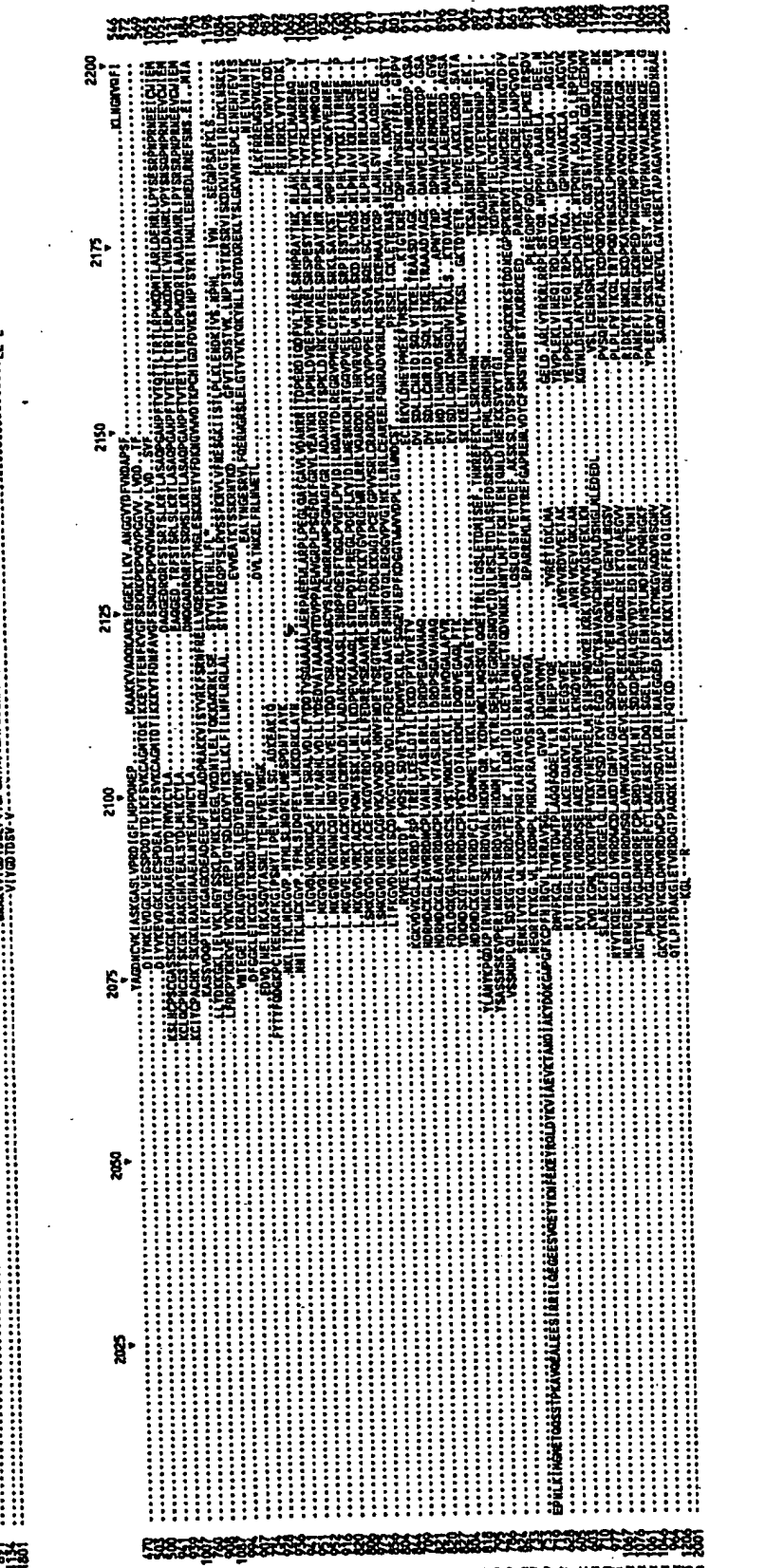


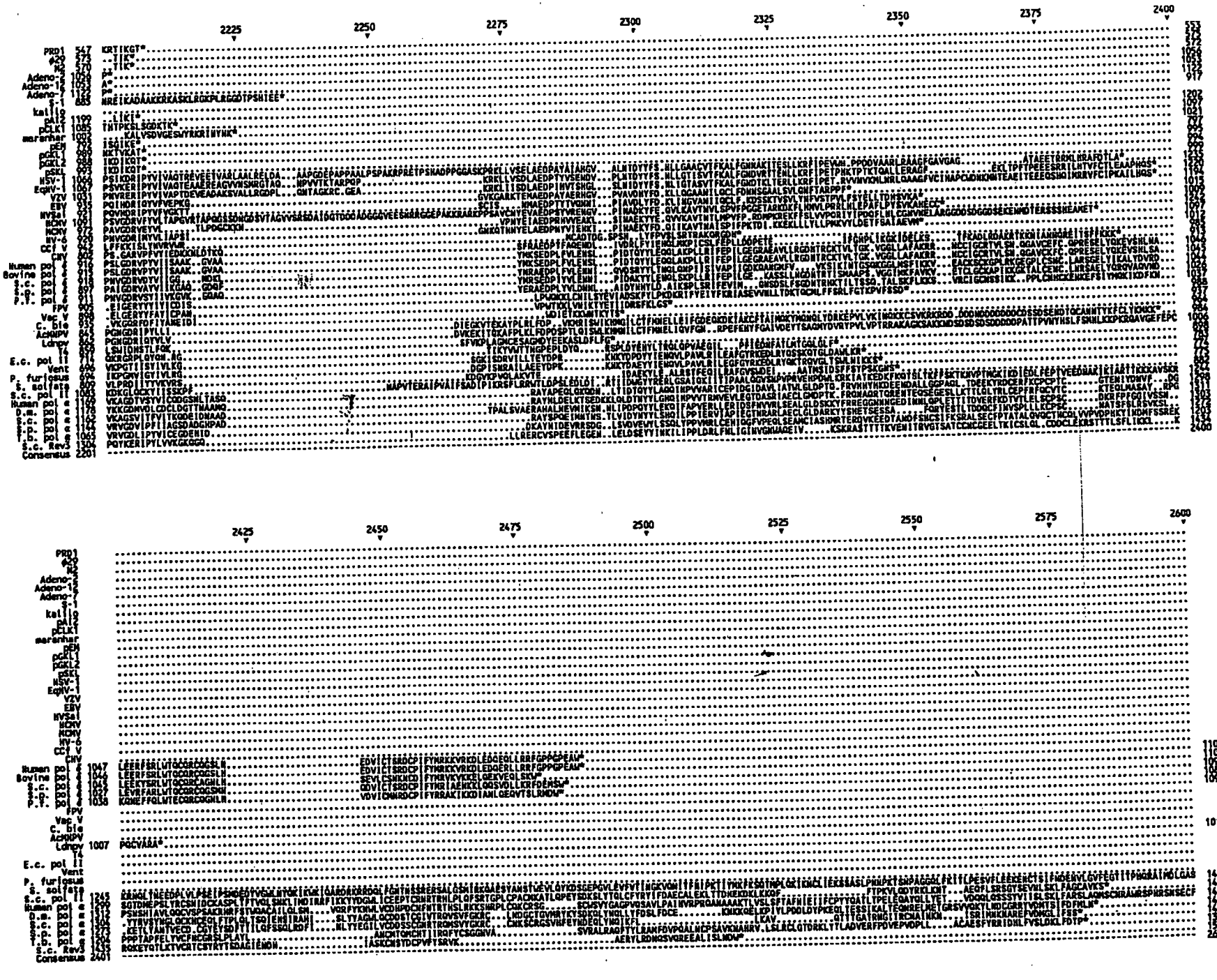


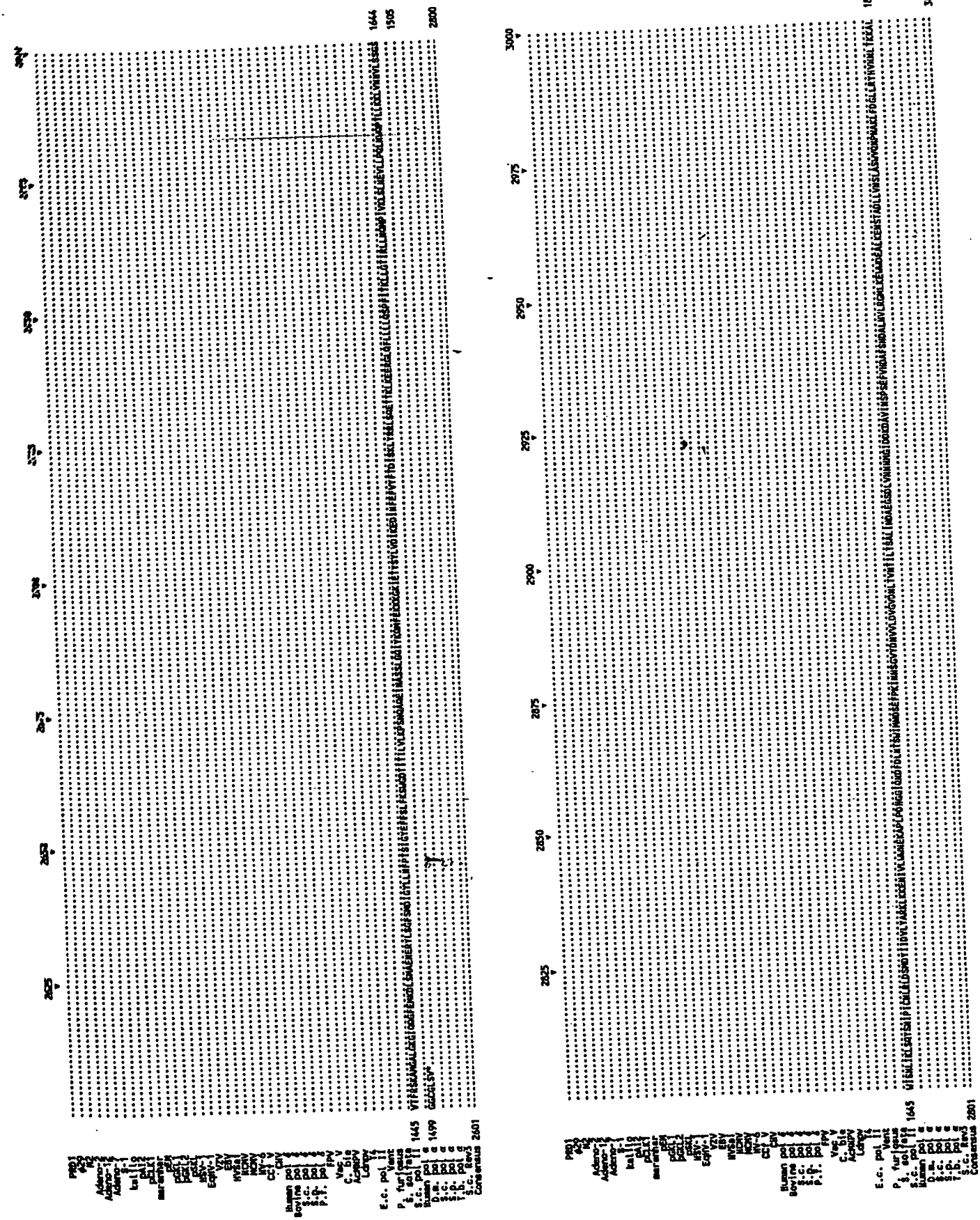


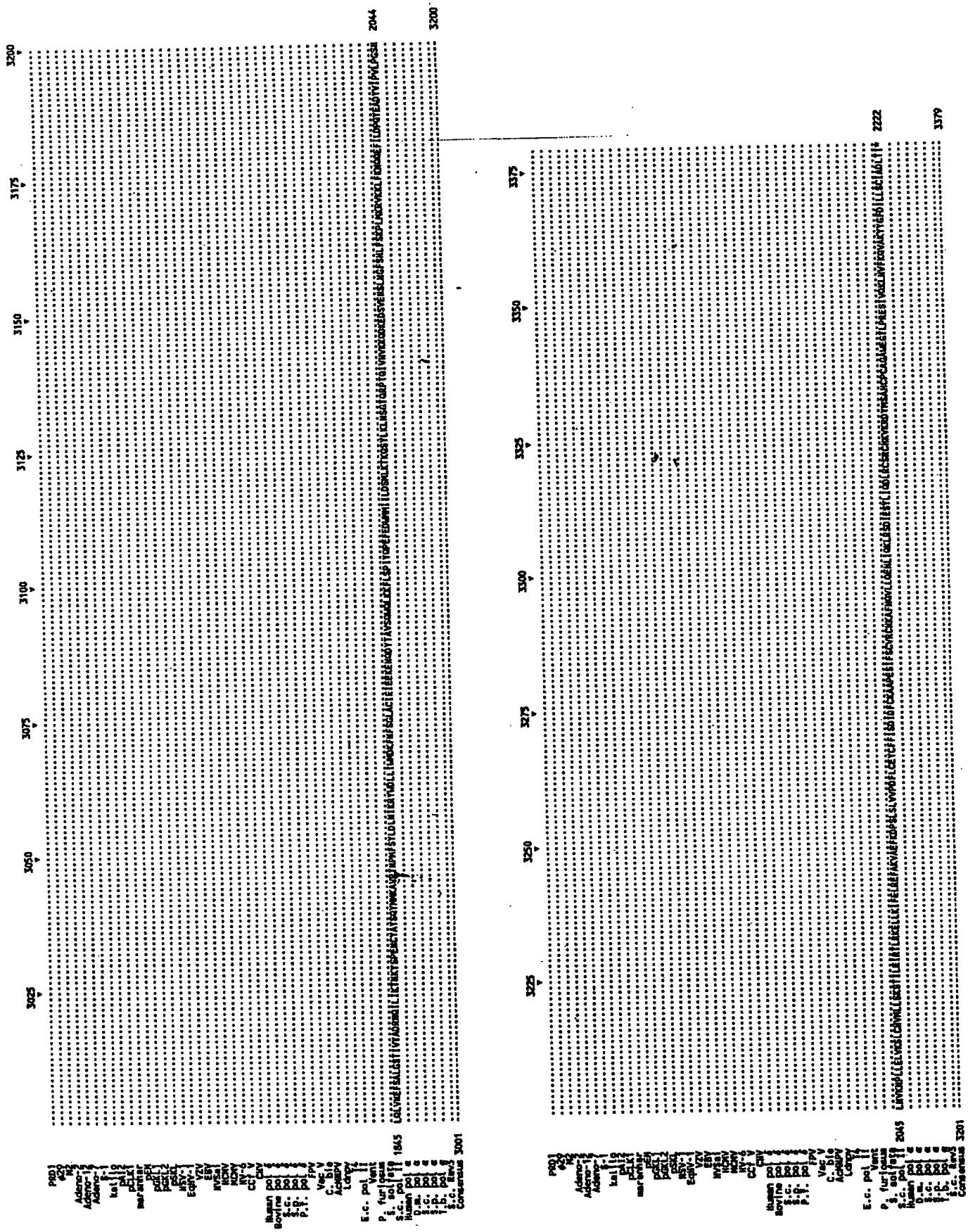












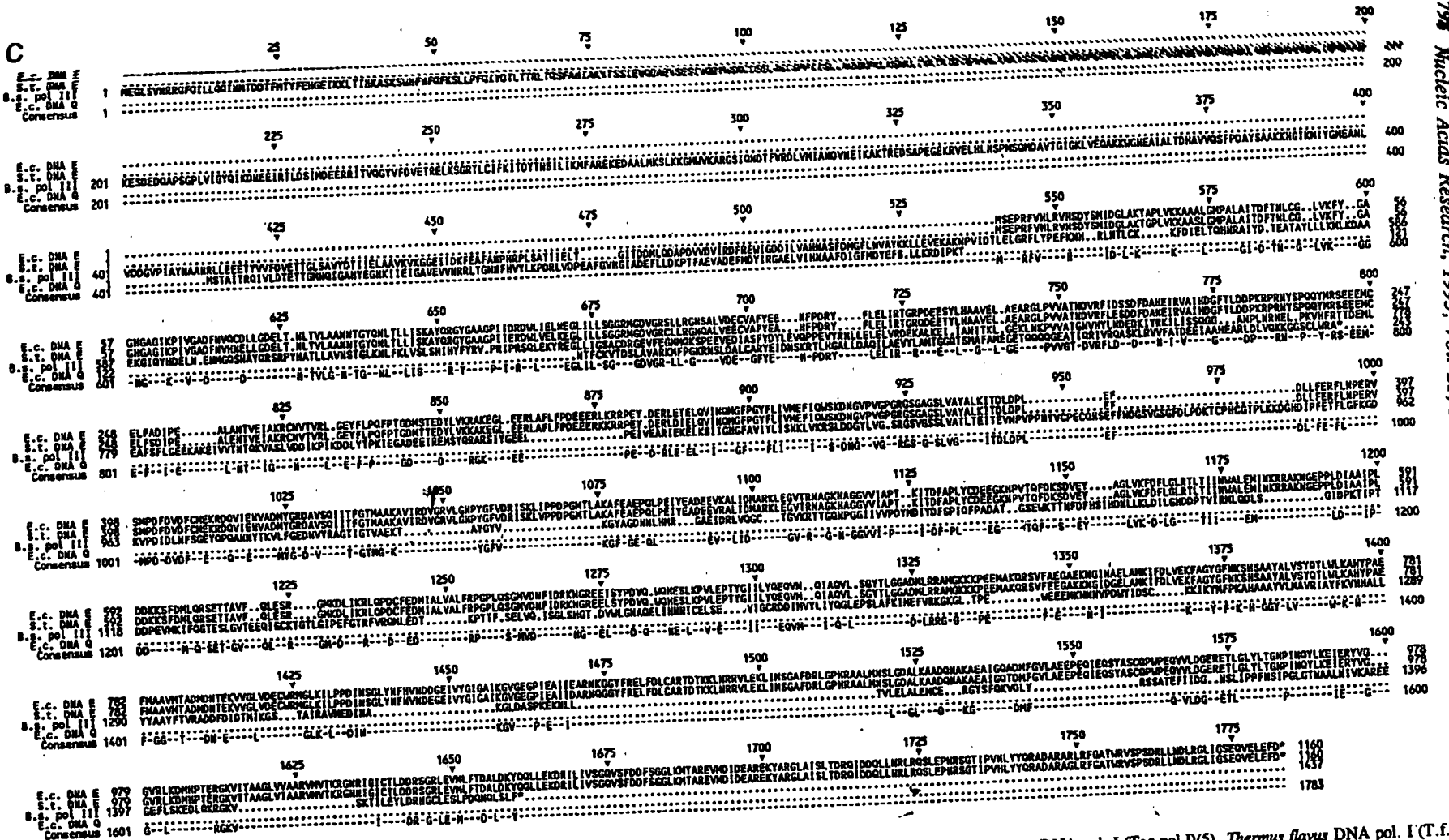


Figure 1A. Family A DNA polymerases—*E.coli* DNA pol. I (E.c. pol. I)(3), *Streptococcus pneumoniae* DNA pol. (S.p. pol I)(4), *Thermus aquaticus* DNA pol. I (Taq pol I)(5), *Thermus flavus* DNA pol. I (T.f. pol I)(6), T5 DNA pol. (7), T7 DNA pol. (8), Spo1 DNA pol. (9), Spo2 DNA pol. (10), yeast mitochondrial DNA pol. (MIP1)(11,12,13), T7 5′-3′ exonuclease (gp 6)(4,8), T5 5′-3′ exonuclease (gp D15)(7,16), and T4 RNase H (gp 33.2)(14,15). B. Family B DNA polymerases—PRD1 DNA pol. (18,19), ϕ29 DNA pol. (20), M2 DNA pol. (21), Adenovirus type-2 DNA pol. (Adeno-2)(54), Adenovirus type-12 DNA pol. (Adeno-12)(56), Adenovirus type-7 DNA pol. (Adeno-7)(55), S-1 maize mitochondrial DNA pol. (S-1)(57), *kalilo neurospora intermedia* DNA pol. (kalilo)(58), *Ascobolus immersus* plasmid DNA pol. (pAI2)(59), *Claviceps purpurea* plasmid DNA pol. (pCLK1)(60), *maranhar neurospora crassa* DNA pol. (maranhar)(61), *Agaricus bitorquis* DNA pol. (pEM)(62), *Kluyveromyces lactis* plasmid DNA pol. (pGKL1)(63), *Kluyveromyces lactis* plasmid DNA pol. (pGKL2)(64), *Saccharomyces Kluyvei* plasmid DNA pol. (pSKL)(65), herpes simplex type-1 DNA pol. (HSV-1)(39), Equine herpes virus type-1 DNA pol. (EqHV-1)(40), Varicella-zoster virus DNA pol. (VZV)(41), Epstein-Barr virus DNA pol. (EBV)(42), *Herpesvirus saimiri* DNA pol. (HVSai)(43), Human cytomegalovirus DNA pol. (HCMV)(44), Murine cytomegalovirus DNA pol. (MCMV)(45), Human herpes virus type-6 DNA pol. (HV-6)(46), Channel Catfish Virus DNA pol. (CCf V)(47), Chlorella virus DNA pol. (NY-2A)(CHV)(48), Human DNA polymerase delta (Human pol δ)(31) (Note: another sequence for Hum. DNA pol delta has been published with 5 amino acid differences(32)), Bovine DNA polymerase delta (Bovine pol δ)(33), *Saccharomyces cerevisiae* DNA polymerase III (delta)(S.c. pol δ)(34),*Schizosaccharomyces pombe* DNA polymerase (delta)(S.p. pol δ)(35), *Plasmodium falciparum* DNA polymerase delta (P.f. pol δ)(36), Fowlpox virus DNA pol. (FPV)(49), Vaccinia virus DNA pol. (Vac V)(50), *Choristoneura biennis* DNA pol. (C. bie)(51), Autographa californica nuclear polyhedrosis virus DNA pol. (AcMNPV)(52), Lymantria dispar nuclear polyhedrosis virus DNA polymerase (LdNPV)(53), T4 DNA pol. (T4)(22), *E.coli* DNA pol. II (E.c. pol II)(17), *Thermococcus litoralis* DNA pol. (Vent)(23), *Pyrococcus furiosus* DNA pol. (P. furious)(24), *Sulfolobus solfataricus* DNA pol. (S. solfata)(25), *Saccharomyces cerevisiae* DNA polymerase II (epsilon)(S.c. pol II)(37), Human DNA polymerase alpha (Human pol α)(26), *Drosophila melanogaster* DNA polymerase alpha (D.m. pol α)(29), *Saccharomyces cerevisiae* DNA polymerase I (alpha)(S.c. pol α)(27), *Schizosaccharomyces Pombe* DNA polymerase I (alpha)(S.p. pol α)(28), *Trypanosoma brucei* DNA polymerase alpha (T.b. pol α)(30), and *Saccharomyces cerevisiae* DNA polymerase Rev3 (S.c. Rev3)(38). C. Family C DNA polymerases—*E.coli* DNA polymerase III α-subunit (E.c. DNA E)(66), *Salmonella typhimurium* DNA polymerase III α-subunit (S.t. DNA E)(67), *Bacillus subtilis* DNA polymerase III (B.S. pol III)(68), and *E.coli* DNA polymerase III ε-subunit (E.c. DNA Q)(69,70).

Table 1. The main families and subclassifications of DNA polymerases. The number of amino acids, molecular weight and isoelectric point of each DNA polymerase or nuclease were computed using the Proteinsort program from the UWGCG. Those Family B DNA polymerases marked with a star (*) are protein-primed DNA polymerases. The sequence marked with (+) pEM (62) is not a complete sequence missing some unknown number of amino acids on the N-terminus and so the numbers in the table only represent this short sequence.

	Amino acids (No.)	Mol. Wt.	Isoelectric pt.	Reference
A. Family A DNA polymerases				
1. Bacterial DNA polymerases				
a) *E.coli* DNA polymerase I	928	103,117	5.37	(3)
b) *Streptococcus pneumoniae* DNA polymerase I	877	99,078	4.78	(4)
c) *Thermus aquaticus* DNA polymerase I	832	93,909	6.38	(5)
d) *Thermus flavus* DNA polymerase I	831	93,783	6.00	(6)
2. Bacteriophage DNA polymerases				
a) T5 DNA polymerase	829	94,410	6.19	(7)
b) T7 DNA polymerase	704	79,691	6.45	(8)
c) Spo1 DNA polymerase	924	106,808	5.34	(9)
d) Spo2 DNA polymerase	648	72,561	8.50	(10)
3. Mitochondrial DNA polymerase				
Yeast mitochondrial DNA polymerase (MIP1)	1254	143,479	9.23	(11,12,13)
4. 5′ to 3′ Exonucleases with homologous sequences to *E.coli* DNA polymerase I				
a) T4 RNase H (gp 33.2)	305	35,558	9.00	(14,15)
b) T5 Exonuclease (gp D15)	291	33,448	5.12	(7,16)
c) T7 Exonuclease (gp 6)	348	40,126	4.54	(4,8)
B. Family B DNA polymerases				
1. Bacterial DNA polymerase				
E.coli DNA polymerase II	783	90,020	6.85	(17)
2. Bacteriophage DNA polymerases				
a) PRD1 DNA polymerase*	553	63,336	6.68	(18,19)
b) φ29 DNA polymerase*	575	66,714	8.83	(20)
c) M2 DNA polymerase*	572	66,423	7.69	(21)
d) T4 DNA polymerase	898	103,609	6.20	(22)
3. Archaebacterial DNA polymerases				
a) *Thermococcus litoralis* DNA polymerase (Vent)	774	89,913	8.29	(23)
b) *Pyrococcus furiosus* DNA polymerase	775	90,112	7.92	(24)
c) *Sulfolobus solfataricus* DNA polymerase	882	101,332	9.72	(25)
4. Eukaryotic Cell DNA polymerases				
(1) DNA polymerase alpha				
a) Human DNA polymerase (alpha)	1,462	165,859	5.71	(26)
b) *S.cerevisiae* DNA polymerase I (alpha)	1,468	166,776	6.14	(27)
c) *S.pombe* DNA polymerase I (alpha)	1,405	159,348	6.85	(28)
d) *Drosophila melanogaster* DNA polymerase (alpha)	1,505	171,167	8.22	(29)
e) *Trypanosoma brucei* DNA polymerase (alpha)	1,339	151,611	6.39	(30)
(2) DNA polymerase delta				
a) Human DNA polymerase (delta)	1,107	123,634	6.94	(31,32)
b) Bovine DNA polymerase (delta)	1,106	123,707	7.52	(33)
c) *S.cerevisiae* DNA polymerase III (delta)	1,097	124,618	7.96	(34)
d) *S.pombe* DNA polymerase III (delta)	1,084	123,211	7.63	(35)
e) *Plasmodium falciparum* DNA polymerase (delta)	1,094	126,883	8.76	(36)
(3) DNA polymerase epsilon				
S.cerevisiae DNA polymerase II (epsilon)	2,222	255,669	6.92	(37)
(4) Other eukaryotic DNA polymerases				
S.cerevisiae DNA polymerase Rev3	1,504	172,956	8.86	(38)
5. Viral DNA polymerases				
a) Herpes Simplex virus type 1 DNA polymerase	1,235	136,547	7.35	(39)
b) Equine herpes virus type 1 DNA polymerase	1,220	135,955	6.67	(40)
c) Varicella-Zoster virus DNA polymerase	1,194	134,047	7.80	(41)
d) Epstein-Barr virus DNA polymerase	1,015	113,417	7.38	(42)
e) *Herpesvirus saimiri* DNA polymerase	1,009	113,934	7.31	(43)
f) Human cytomegalovirus DNA polymerase	1,242	137,101	7.25	(44)
g) Murine cytomegalovirus DNA polymerase	1,097	123,573	6.68	(45)
h) Human herpes virus type 6 DNA polymerase	1,012	115,819	7.11	(46)
i) Channel Catfish virus DNA polymerase	985	113,468	7.98	(47)
j) Chlorella virus DNA polymerase	913	104,955	6.66	(48)
k) Fowlpox virus DNA polymerase	988	116,658	8.11	(49)
l) Vaccinia virus DNA polymerase	937	108,564	7.50	(50)
m) Choristoneura biennis DNA polymerase	964	114,818	7.95	(51)
n) Autographa californica nuclear polyhedrosis virus (AcMNPV) DNA polymerase	984	114,337	8.35	(52)
o) Lymantria dispar nuclear polyhedrosis virus DNA polymerase	1,013	115,921	9.08	(53)

p) Adenovirus-2 DNA polymerase*	1,056	120,431	6.65	(54)
q) Adenovirus-7 DNA polymerase*	1,122	128,648	6.73	(55)
r) Adenovirus-12 DNA polymerase*	1,053	120,863	6.86	(56)
6. Eukaryotic linear DNA plasmid encoded DNA polymerases				
a) S-1 maize DNA polymerase*	917	105,935	8.62	(57)
b) *kalilo neurospora intermedia* DNA polymerase*	970	112,902	9.71	(58)
c) pAI2 *Ascobolus immersus* DNA polymerase*	1,202	138,279	10.10	(59)
d) pCLK1 *Claviceps purpurea* DNA polymerase*	1,097	126,627	8.76	(60)
e) *maranhar neurospora crassa* DNA polymerase*	1,021	119,074	9.62	(61)
f) pEM *Agaricus bitorquis* DNA polymerase*	†797	†91,922	†8.24	(62)
g) pGKL1 *Kluyveromyces lactis* DNA polymerase*	995	116,345	8.04	(63)
h) pGKL2 *Kluyveromyces lactis* DNA polymerase*	994	117,560	8.33	(64)
i) pSKL *Saccharomyces kluyveri* DNA polymerase*	999	117,544	9.79	(65)
C. Family C DNA polymerases				
1. Bacterial replicative DNA polymerases				
a) *E.coli* DNA polymerase III α subunit	1160	129,903	5.04	(66)
b) *S. typhimurium* DNA polymerase III α subunit	1160	130,118	5.05	(67)
c) *Bacillus subtilis* DNA polymerase III	1437	162,648	5.23	(68)
2. *E.coli dnaQ* (MutD)				
E.coli DNA polymerase III ϵ subunit	243	27,099	5.68	(69,70)

4. Lopez, P., Martinez, S., Diaz, A., Espinosa, M. and Lacks, S. A. (1989) J. Biol. Chem. **264**, 4255–4263.
5. Lawyer, F. C., Stoffel, S., Saiki, R. K., Myambo, K., Drummond, R. and Gelfand, D. H. (1989) J. Biol. Chem. **264**, 6427–6437.
6. Akhmetzjanov, A.A. and Vakhitov, V.A. (1992) Nucl. Acids Res. **20**, 5839.
7. Leavitt, M. C. and Ito, J. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 4465–4469.
8. Dunn, J. J. and Studier, F. W. (1983) J. Mol. Biol. **166**, 477–535.
9. Scarlato, V. and Gargano, S. (1992) Gene **118**, 109–113.
10. Rådén, B. and Rutberg, L. (1984) J. Virol. **52**, 9–15.
11. Foury, F. (1989) J. Biol. Chem. **264**, 20552–20560.
12. Ito, J. and Braithwaite, D.K. (1990) Nucl. Acids Res. **18**, 6716.
13. Blanco, L., Bernad, A. and Salas, M. (1991) Nucl. Acids Res. **19**, 955.
14. Hahn, S. and Rüger, W. (1989) Nucl. Acids Res. **17**, 6729.
15. Hollingsworth, H.C. and Nossal, N.G. (1991) J. Biol. Chem. **266**, 1888–1897.
16. Kaliman, A.V., Krutilina, A.I., Kryukov, V.M. and Bayev, A.A. (1986) FEBS Lett. **195**, 61–64.
17. Iwasaki, H., Ishino, Y., Toh, H., Nakata, A. and Shinagawa, H. (1991) Mol. Gen. Genet. **226**, 24–33.
18. Jung, G., Leavitt, M. C., Hsieh, J.-C. and Ito, J. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 8287–8291.
19. Savilahti, H. and Bamford D.H. (1987) Gene **57**, 121–130.
20. Yoshikawa, H. and Ito, J. (1982) Gene **17**, 323–335.
21. Matsumoto, K., Takano, H., Kim, C. I. and Hirokawa, H. (1989) Gene **84**, 247–255.
22. Spicer, E. K., Rush, J., Fung, C., Reha-Krantz, L. J., Karam, J. D. and Konigsberg, W. H. (1988) J. Biol. Chem. **263**, 7478–7486.
23. Perler, F.B., Comb, D.G., Jack, W.E., Moran, L.S., Qiang, B., Kucera, R.B., Benner, J., Slatko, B.E., Nwankwo, D.O., Hempstead, S.K., Carlow, C.K.S. and Jannasch, H. (1992) Proc. Natl. Acad. Sci. USA **89**, 5577–5581.
24. Cline, J.M., Scott, B.R., Adams, M.W.W., Sorge, J.A. and Mathur, E.J. Gene (in press).
25. Pisani, F.M., De Martino, C. and Rossi, M. (1992) Nucl. Acids Res. **20**, 2711–2716.
26. Wong, S. W., Wahl, A. F., Yuan, P.-M., Arai, N., Pearson, B. E., Arai, K.-i., Korn, D., Hunkapiller, M. W. and Wang, T.S.-F. (1988) EMBO J. **7**, 37–47.
27. Pizzagalli, A., Valsasnini, P., Plevani, P. and Lucchini, G. (1988)Proc. Natl. Acad. Sci. U.S.A. **85**, 3772–3776.
28. Damagnez, V., Tillit, J., de Recondo, A.-M. and Baldacci, G. (1991) Mol. Gen. Genet. **226**, 182–189.
29. Hirose, F., Yamaguchi, M., Nishida, Y., Masutani, M., Miyazawa, H., Hanaoka, F. and Matsukage, A. (1991) Nucl. Acids Res. **19**, 4991–4998.
30. Leegwater, P.A.J., Strating, M., Murphy, N.B., Kooy, R.F., van der Vliet, P.C. and Overdulve, J.P. (1991) Nucl. Acids Res. **19**, 6441–6447.
31. Chung, D.W., Zhang, J., Tan C.-K., Davie, E.W., So, A.G. and Downey, K.M. (1991) Proc. Natl. Acad. Sci. USA **88**, 11197–11201.
32. Yang, C.-L., Chang, L.-S., Zhang, P., Hao, H., Zhu, L., Tommey, N.L. and Lee, M.Y.W.T. (1992) Nucl. Acids Res. **20**, 735–745.
33. Zhang, J., Chung, D.W., Tan, C.-K., Downey, K.M., Davie, E.W. and So, A.G. (1991) Biochemistry **30**, 11742–11750.
34. Morrison, A. and Sugino, A. (1992) Nucl. Acids Res. **20**, 375.
35. Pignéde, G., Bouvier, D., de Recondo, A.-M. and Baldacci, G. (1991) J. Mol. Biol. **222**, 209–218.
36. Ridley, R.G., White, J.H., McAleese, S.M., Goman, M., Alano, P., de Vries, E. and Kilbey, B.J. (1991) Nucl. Acids Res. **19**, 6731–6736.
37. Morrison, A., Araki, H., Clark, A.B., Hamatake, R.K. and Sugino, A. (1990) Cell **62**, 1143–1151.
38. Morrison, A., Christensen, R. B., Alley, J., Beck, A. K., Bernstine, E. G., Lemontt, J. F. and Lawrence, C. W. (1989) J. Bacteriol. **171**, 5659–5667.
39. Gibbs, J. S., Chiou, H. C., Hall, J. D., Mount, D. W., Retondo, M. J., Weller, S. K. and Coen, D. M. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 7969–7973.
40. Telford, E.A., Watson, M.S., McBride, K. and Davison, A.J. (1992) Virology **189**, 304–316.
41. Davison, A. J. and Scott, J. E. (1986) J. Gen. Virol. **67**, 1759–1816.
42. Baer, R., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G., Hudson, G.S., Satchwell, S.C., Séguin, C., Tuffnell, P.S. and Barrell, B.G. (1984) Nature **310**, 207–211.
43. Albrecht, J.-C. and Fleckenstein, B. (1990) Virology **174**, 533–542.
44. Kouzarides, T., Bankier, A. T., Satchwell, S. C., Weston, K., Tomlinson, P. and Barrell, B. G. (1987) J. Virol. **61**, 125–133.
45. Elliott, R., Clark, C. Jaquish, D. and Spector, D.H. (1991) Virology **185**, 169–186.
46. Teo, I.A., Griffin, B.E. and Jones, M.D. (1991) J. Virol. **65**, 4670–4680.
47. Davison, A.J. (1992) Virology **186**, 9–14.
48. Grabherr, R., Strasser, P. and Van Etten, J.L. (1992) Virology **188**, 721–731.
49. Binns, M. M., Stenzler, L., Tomley, F. M., Campbell, J. and Boursnell, M. E. G. (1987) Nucl. Acids Res. **15**, 6563–6573.
50. Earl, P. L., Jones, E. V. and Moss, B. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 3659–3663.
51. Mustafa, A. and Yuen, L. (1991) DNA Seq. **2**, 39–45.
52. Tomalski, M. D., Wu, J. and Miller, L. K. (1988) Virology **167**, 591–600.
53. Bjornson, R.M. and Rohrmann, G.F. (1992) J. Gen. Virol. **73**, 1499–1504.
54. Gingeras, T. R., Sciaky, D., Gelinas, R. E., Bing-Dong, J., Yen, C. E., Kelly, M. M., Bullock, P. A., Parsons, B. L., O'Neill, K. E. and Roberts, R. J. (1982) J. Biol. Chem. **257**, 13475–13491.
55. Engler, J.A., Hoppe, M.S. and van Bree, M.P. (1983) Gene **21**, 145–159.
56. Shu, L., Hong, J.S., Wei, Y.-f. and Engler, J.A. (1986) Gene **46**, 187–195.
57. Paillard, M., Sederoff, R. R. and Levings, C. S. III (1985) EMBO J. **4**, 1125–1128.
58. Chan, B.S.-S., Court, D.A., Vierula, P.J. and Bertrand, H. (1991) Curr. Genet. **20**, 225–237.

59. Kempken, F., Meinhardt, F. and Esser, K. (1989) Mol. Gen. Genet. **218**, 523–530.
60. Oeser, B. and Tudzynski, P. (1989) Mol. Gen. Genet. **217**, 132–140.
61. Court D.A. and Bertrand, H. (1992) Curr. Genet. **22**, 385–397.
62. Robison, M.M., Royer, J.C. and Horgen, P.A. (1991) Curr. Genet. **19**, 495–502.
63. Stark, M.J.R., Mileham, A.J., Romanos, M.A. and Boyd, A. (1984) Nucl. Acids Res. **12**, 6011–6030.
64. Tommasino, M., Ricci, S. and Galeotti, C. L. (1988) Nucl. Acids Res. **16**, 5863–5878.
65. Hishimuma, F. and Hirai, K. (1991) J. Gen. Genet. **226**, 97–106.
66. Tomasiewicz, H. G. and McHenry, C. S. (1987) J. Bacteriol. **169**, 5735–5744.
67. Lancy, E. D., Lifsics, M. R., Munson, P. and Maurer, R. (1989) J. Bacteriol. **171**, 5581–5586.
68. Hammond, R. A., Barnes, M. H., Mack, S. L., Mitchener, J. A. and Brown, N. C. (1991) Gene **98**, 29–36.
69. Maki, H., Horiuchi, T. and Sekiguchi, M. (1983) Proc. Natl. Acad. Sci. USA **80**, 7137–7141.
70. Joyce, C.M., Ollis, D.L., Rush, J., Steitz, T.A., Konigsberg, W.H. and Grindley, N.D.F. (1986) in 'Protein structure, folding and design (UCLA Symposia on Molecular and Cellular Biology, Vol 32)', D. Oxender, Ed., pp 197–205, Alan R. Liss.
71. Felsenstein, J. (1989) Cladistics **5**, 164–166.

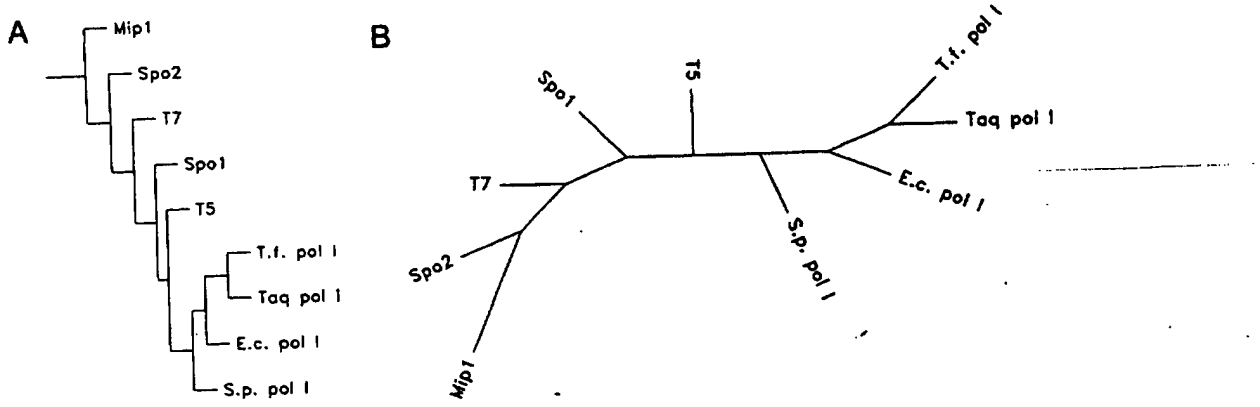


Figure 2A. Phylogenetic phenogram tree produced from the alignment of Family A DNA polymerases in Figure 1A using only the following conserved sequence blocks of the DNA polymerase domain: 798–814, 877–998, 1047–1090, 1104–1123, 1131–1158, 1175–1206, 1236–1251, 1284–1305, 1322–1340, and 1365–1379. **B.** Unrooted phylogenetic tree of the Family A DNA polymerases produced as in Figure 2A plotted by a different method.

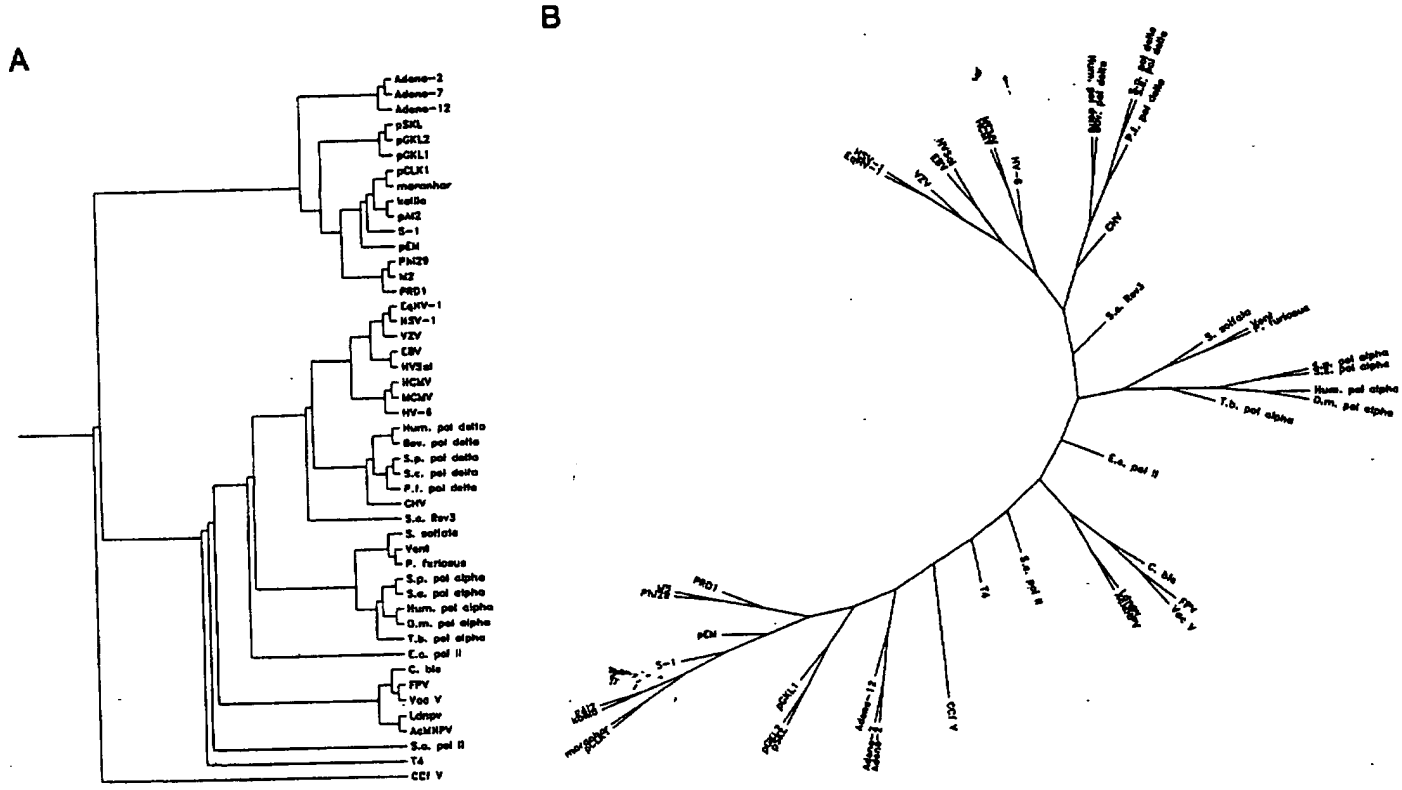


Figure 3A. Phylogenetic phenogram tree produced from the alignment of Family B DNA polymerases in Figure 1B using only the following conserved sequence blocks of the DNA polymerase domain: 1407–1760, 1885–1901, 1956–1990, 2081–2100, 2181–2210, and 2280–2320. **B.** Unrooted phylogenetic tree of the Family B DNA polymerases produced as in Figure 3A plotted by a different method.

ACKNOWLEDGEMENTS

The authors are grateful to all the individuals who have kindly communicated their unpublished sequences to be included in this compilation. This work was supported in part by Grant GM28013 from the NIH to J.I.

REFERENCES

1. Ito, J. and Braithwaite, D.K. (1991) Nucl. Acids Res. 19, 4045–4057.
2. Wheelis, M.L., Kandler, O. and Woese, C.R. (1992) Proc. Natl. Acad. Sci. USA 89, 2930–2934.
3. Joyce, C. M., Kelley, W. S. and Grindley N. D. F. (1982) J. Biol. Chem. 257, 1958–1964.

Nucleic Acids Research, Vol. 19, No. 15 4045–4057

Compilation and alignment of DNA polymerase sequences

Junetsu Ito and Dan K.Braithwaite
Department of Microbiology and Immunology, College of Medicine, University of Arizona Health Sciences Center, Tucson, AZ 85724, USA

Received May 29, 1991; Revised and Accepted July 9, 1991

INTRODUCTION

More than 40 different DNA polymerases, including some putative DNA polymerase sequences deduced from nucleotide sequence data, have recently been reported (1–39). The amino acid sequences of these DNA polymerases have been aligned and partial homologous regions identified by many investigators (2–4,9,10,12–25,27–36,42–51). Based on the segmental amino acid sequence similarities, DNA polymerases have been classified into two major groups; *E. coli* DNA polymerase I-Type and eukaryotic DNA polymerase α-Type (14,44,47,48,51), or family A DNA polymerases and family B DNA polymerases (4,9,50). As the number of DNA polymerase sequences increases, the classification of DNA polymerases becomes increasingly ambiguous. For example, DNA polymerase delta of yeast was shown to have amino acid sequence similarity to the α-Type DNA polymerases (17). It has become necessary to establish a unified classification of DNA polymerases. Here we pr pose to classify DNA polymerases into families A, B, and C (Figure 1: A, B, and C), according to the amino acid sequence homologies with *E. coli* DNA polymerases I, II, and III, respectively. As new and different prokaryotic and eukaryotic DNA polymerases are identified, the number of families can easily be expanded by using additional letters of the alphabet (i.e., D, E, etc.).

The bacterium *E. coli* (strain K12) contains three distinct DNA polymerases I, II, and III (52). *E. coli* DNA polymerase I, the first DNA polymerase discovered, is specified by the *polA* gene (52). *E. coli* DNA polymerase II, encoded by the *polB* gene, was recently sequenced and found to be identical to the *din*A gene, a DNA damage inducible gene whose expression is regulated by the SOS system in *E. coli* (8,53). Amino acid sequence alignment shows that *E. coli* DNA polymerase II has significant homology with family B (α-Type) DNA polymerases (8,53,54).

E. coli DNA polymerase III is a multisubunit enzyme encoded by various *dna* genes (55); the DNA polymerizing α-subunit encoded by the *polC* (*dnaE*) gene (56) and the 3′→5′ exonuclease performing ε-subunit encoded by the *dnaQ* gene (57). The α-subunit of *E. coli* DNA polymerase III exhibits an extensive hom logy with the corresponding α-subunit of *Salmonella typhimurium* DNA polymerase III (35); and both show significant homology to *Bacillus subtilis* DNA polymerase III, a single-polypeptide encoded by the *polC* gene (36).

In summary, family A DNA polymerases are named for their homology to the product of the *polA* gene encoding *E. coli* DNA polymerase I; family B DNA polymerases are named for their homology to the product of the *polB* gene encoding *E. coli* DNA polymerase II; and family C DNA polymerases are named for their homology to the product of the *polC* gene encoding *E. coli* DNA polymerase III.

The eukaryotic DNA polymerase β, the smallest known DNA polymerase, does not have homology with those of any of the DNA polymerase families described above. Instead, DNA polymerase β has homology with terminal transferases (37). This β group we will call family X (Figure 1D). The classification and original reference(s) for the amino acid sequences of each DNA polymerase are shown in Table 1.

All of the family A DNA polymerases, except for yeast mitochondrial DNA polymerase I, are prokaryotic and are very sensitive to dideoxynucleotide inhibitors, and therefore are useful enzymes for DNA sequencing by the chain-termination method (58). The family A DNA polymerases are resistant to aphidicolin. The family B DNA polymerases are quite extensive in number and variety. Most of the family B DNA polymerases, if not all, are sensitive to aphidicolin and relatively resistant to dideoxynucleotide inhibitors. Most of the family B DNA polymerases, except for pAI2 (33) and yeast DNA polymerase II (16), contain the highly conserved amino acid sequence motif **YGDTD**, which has been suggested to form part of the dNTP binding site. Amino acid substitutions in this conserved sequence resulted in defects in the DNA polymerase activity without affecting the 3′→5′ exonuclease activity (59,60,61). The family C DNA polymerases are major bacterial replicative DNA polymerases which do not have appreciable homology with those of family A and B DNA polymerases. *B. subtilis* DNA polymerase III is a single polypeptide that is highly sensitive to hydroxyphenylazouracil (62). It is anticipated that the number of sequenced family C DNA polymerases will increase rapidly, since all of the aerobic bacteria may contain a member of this family of DNA polymerases.

SEQUENCE ALIGNMENT

The 37 complete DNA polymerase sequences and 3 complete terminal deoxynucleotidyltransferase (TDT) sequences are listed in 4 groups; the family A DNA polymerases, the family B DNA polymerases, the family C DNA polymerases, and family X DNA polymerases (including TDTs). In order to limit the space needed for the alignment, we omitted DNA polymerase sequences that are very similar to th prototype DNA polymerase. The DNA polymerases not shown include: herpes virus type-2 (63),

adenovirus type-5 (64), bacteriophage T3 (65), and bacteriophage PZA (66).

ACCURACY OF SEQUENCE DATA

Whenever a sequence ambiguity existed in a published sequence, we contacted the authors to obtain the updated sequence information. We found that a few published amino acid sequences differ at one or more positions from their GenBank/EMBL entry. Again, we have communicated with the primary author to confirm the correct sequences.

The multiple alignment of the amino acid sequences was obtained by a series of pairwise alignments combined and adjusted by eye into larger and larger subsets of similar sequences. The process of combining and adjusting by eye was aided by modified versions of the MOTIF program (67) and the ALIGN program (68). The GAP and BESTFIT programs, from UWGCG (University of Wisconsin Genetic Computer Group) (69), initially generated the pairwise alignments, adjusted for maximum alignment that allowed for a considerable number of gaps. We then compressed these alignments by eye to give a more contiguous alignment. The alignment of the sequences for optimal similarity is straightforward in the areas of relatively conserved structure but is much more arbitrary in the more varied sequence areas. The alignment of the varied areas should therefore be regarded as less than optimal in view of the difficulties concerned with multiple alignments in these areas.

Finally, we invite further correction from readers, and welcome suggested revisions and alternative alignments.

ACKNOWLEDGEMENTS

This work was supported by grant GM28013 from the National Institutes of Health and by grant NP-704 from the American Cancer Society.

REFERENCES

1. Joyce, C.M., Kelley, W.S. and Grindley N.D.F. (1982) *J. Biol. Chem.* **257**, 1958–1964.
2. Lopez, P., Martinez, S., Diaz, A., Espinosa, M. and Lacks, S.A. (1989) *J. Biol. Chem.* **264**, 4255–4263.
3. Lawyer, F.C., Stoffel, S., Saiki, R.K., Myambo, K., Drummond, R. and Gelfand, D.H. (1989) *J. Biol. Chem.* **264**, 6427–6437.
4. Leavitt, M.C. and Ito, J. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4465–4469.
5. Dunn, J.J. and Studier, F.W. (1983) *J. Mol. Biol.* **166**, 477–535.
6. Rådén, B. and Rutberg, L. (1984) *J. Virol.* **52**, 9–15.
7. Four, F. (1989) *J. Biol. Chem.* **264**, 20552–20560.
8. Iwasaki, H., Ishino, Y., Toh, H., Nakata, A. and Shinagawa, H. (1991) *Mol. Gen. Genet.* **226**, 24–33.
9. Jung, J., Leavitt, M.C., Hsieh, J.-C. and Ito, J. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8287–8291.
10. Savilahti, H. and Bamford D.H. (1987) *Gene* **57**, 121–130.
11. Yoshikawa, H. and Ito, J. (1982) *Gene* **17**, 323–335.
12. Matsumoto, K., Takano, H., Kim, C.I. and Hirokawa, H. (1989) *Gene* **84**, 247–255.
13. Spicer, E.K., Rush, J., Fung, C., Reha-Krantz, L.J., Karam, J.D. and Konigsberg, W.H. (1988) *J. Biol. Chem.* **263**, 7478–7486.
14. Wong, S.W., Wahl, A.F., Yuan, P.-M., Arai, N., Pearson, B.E., Arai, K.-i., Korn, D., Hunkapiller, M.W. and Wang, T.S.-F. (1988) *EMBO J.* **7**, 37–47.
15. Pizzagalli, A., Valsasnini, P., Plevani, P. and Lucchini, G. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3772–3776.
16. Morrison, A., Araki, H., Clark, A.B., Hamatake, R.K. and Sugino, A. (1990) *Cell* **62**, 1143–1151.
17. Boulet, A., Simon, M., Faye, G., Bauer, G.A. and Burgers, P.M.J. (1989) *EMBO J.* **8**, 1849–1854.
18. Morrison, A., Christensen, R.B., Alley, J., Beck, A.K., Bernstine, E.G., Lemontt, J.F. and Lawrence, C.W. (1989) *J. Bacteriol.* **171**, 5659–5667.
19. Gibbs, J.S., Chiou, H.C., Hall, J.D., Mount, D.W., Retondo, M.J., Weller, S.K. and Coen, D.M. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7969–7973.
20. Kouzarides, T., Bankier, A.T., Satchwell, S.C., Weston, K., Tomlinson, P. and Barrell, B.G. (1987) *J. Virol.* **61**, 125–133.
21. Baer, R., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G., Hudson, G.S., Satchwell, S.C., Séguin, C., Tuffnell, P.S. and Barrell, B.G. (1984) *Nature* **310**, 207–211.
22. Davison, A.J. and Scott, J.E. (1986) *J. Gen. Virol.* **67**, 1759–1816.
23. Binns, M.M., Stenzler, L., Tomley, F.M., Campbell, J. and Boursnell, M.E.G. (1987) *Nucl. Acids Res.* **15**, 6563–6573.
24. Earl, P.L., Jones, E.V. and Moss, B. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3659–3663.
25. Tomalski, M.D., Wu, J. and Miller, L.K. (1988) *Virology* **167**, 591–600.
26. Gingeras, T.R., Sciaky, D., Gelinas, R.E., Bing-Dong, J., Yen, C.E., Kelly, M.M., Bullock, P.A., Parsons, B.L., O'Neill, K.E. and Roberts, R.J. (1982) *J. Biol. Chem.* **257**, 13475–13491.
27. Engler, J.A., Hoppe, M.S. and van Bree, M.P. (1983) *Gene* **21**, 145–159.
28. Shu, L., Hong, J.S., Wei, Y.-f. and Engler, J.A. (1986) *Gene* **46**, 187–195.
29. Paillard, M., Sederoff, R.R. and Levings, C.S. III (1985) *EMBO J.* **4**, 1125–1128.
30. Stark, M.J.R., Mileham, A.J., Romanos, M.A. and Boyd, A. (1984) *Nucl. Acids Res.* **12**, 6011–6030.
31. Tommasino, M., Ricci, S. and Galeotti, C.L. (1988) *Nucleic Acids Res.* **16**, 5863–5878.
32. Oeser, B. and Tudzynski, P. (1989) *Mol. Gen. Genet.* **217**, 132–140.
33. Kempken, F., Meinhardt, F. and Esser, K. (1989) *Mol. Gen. Genet.* **218**, 523–530.
34. Tomasiewicz, H.G. and McHenry, C.S. (1987) *J. Bacteriol.* **169**, 5735–5744.
35. Lancy, E.D., Lifsics, M.R., Munson, P. and Maurer, R. (1989) *J. Bacteriol.* **171**, 5581–5586.
36. Hammond, R.A., Barnes, M.H., Mack, S.L., Mitchener, J.A. and Brown, N.C. (1991) *Gene* **98**, 29–36.
37. Matsukage, A., Nishikawa, K., Ooi, T., Seto, Y. and Yamaguchi, M. (1987) *J. Biol. Chem.* **262**, 8960–8962.
38. Abbotts, J., SenGupta, D.N., Zmudzka, B., Widen, S.G., Notario, V. and Wilson, S.H. (1988) *Biochemistry* **27**, 901–909.
39. SenGupta, D.N., Zmudzka, B.Z., Kumar, P., Cobianchi, F., Skowronski, J. and Wilson, S.H. (1986) *Biochem. Biophys. Res. Comm.* **136**, 341–347.
40. Peterson, R.C., Cheung, L.C., Mattaliano, R.J., White, S.T., Chang, L.M.S. and Bollum F.J. (1985) *J. Biol. Chem.* **260**, 10495–19502.
41. Koiwai, O., Yokota, T., Kageyama, T., Hirose, T., Yoshida, S. and Arai, K.-i. (1986) *Nucl. Acids Res.* **14**, 5777–5792.
42. Argos, P., Tucker, A.D. and Phillipson, L. (1986) *Virology* **149**, 208–216.
43. Larder, B.A., Kemp, S.D. and Darby, G. (1987) *EMBO J.* **6**, 160–175.
44. Hall, J.D. (1988) *Trends Genet.* **4**, 42–46.
45. Reha-Krantz, L.J. (1988) *J. Mol. Biol.* **202**, 711–724.
46. Bernad, A., Zaballos, A., Salas, M. and Blanco, L. (1987) *EMBO J.* **6**, 4219–4225.
47. Wang, T.S.-F., Wong, S.W. and Korn, D. (1989) *FASEB* **3**, 14–21.
48. Bernad, A., Blanco, L., Lazaro, J.M., Martin, G. and Salas, M. (1989) *Cell* **59**, 219–228.
49. Polesky, A.H., Steitz, T.A., Grindley, N.D.F. and Joyce, C.M. (1990) *J. Biol. Chem.* **24**, 14579–14591.
50. Ito, J. and Braithwaite, D.K. (1990) *Nucl. Acids Res.* **18**, 6716.
51. Blanco, L., Bernad, A. and Salas, M. (1991) *Nucl. Acids Res.* **19**, 955.
52. Kornberg, A. (1974) DNA replication. W.H. Freeman and Co., San Francisco.
53. Bonner, C.A., Hays, S., McEntee, K. and Goodman M.F. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 7663–7667.
54. Chen, H., Lawrence, C.B., Bryan, S.K. and Moses, R.E. (1990) *Nucl. Acids Res.* **18**, 7185–7186.
55. McHenry, C.S.(1985) *Mol. Cell. Biochem.* **66**, 71–85.
56. Shepard, D., Oberfelder, R.W., Welch, M.W. and McHenry, C.S. (1984) *J. Bacteriol.* **158**, 455–459.
57. Scheuermann, R., Tam, S., Burgers, P.M.J., Lu, C. and Echols, H. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 7085–7089.
58. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467.
59. Dorsky, D.I. and Crumpacker, C.S. (1990) *J. Virol.* **64**, 1394–1397.
60. Bernad, A., Lázaro, J.M., Salas, M. and Blanco, L. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4610–4614.

61. Jung, G., Leavitt, M.C., Schultz, M. and Ito, J. (1990) *Biochem. Biophys. Res. Comm.* 170, 1294–1300.
62. Neville, M.N. and Brown, N.C. (1972) *Nature New Biol.* 240, 80–82.
63. Tsurumi, T., Maeno, K. and Nishiyama, Y. (1987) *Gene* 52, 129–137.
64. Dekker, B.M.M. and Van Ormondt, H. (1984) *Gene* 27, 115–120.
65. Beck, P.J., Gonzalez, S., Ward, C.L. and Molineux, I.J. (1989) *J. Mol. Biol.* 210, 687–701.
66. Paces, V., Vlcek, C., Urbanek, P. and Hostomsky, Z. (1985) *Gene* 38, 45–56.
67. Smith, H.O., Annau, T.M. and Chandrasegaran, S. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 826–830.
68. Doolittle, R.F.and Feng, D.-F. (1990) In Doolittle, R.F. (ed.), Methods in Enzymol.–Molecular Evolution: Computer Analysis of Protein and Nucleic Acid Sequences. Academic Press, New York, Vol. 183, pp. 659–669.
69. Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucl. Acids Res.* 12, 387–395.

Classification of DNA polymerases

		References
A.	**Family A DNA polymerases**	
	1. Bacterial DNA polymerases	
	a) *E. coli* DNA polymerase I	(1)
	b) *Streptococcus pneumoniae* DNA polymerase I	(2)
	c) *Thermus aquaticus* DNA polymerase I	(3)
	2. Bacteriophage DNA polymerases	
	a) T5 DNA polymerase	(4)
	b) T7 DNA polymerase	(5)
	c) Spo2 DNA polymerase	(6)
	3. Mitochondrial DNA polymerase	
	Yeast mitochondrial DNA polymerase (MIP1)	(7)
B.	**Family B DNA polymerases**	
	1. Bacterial DNA polymerases	
	E. coli DNA polymerase II	(8)
	2. Bacteriophage DNA polymerases	
	a) PRD1 DNA polymerase*	(9,10)
	b) ϕ29 DNA polymerase*	(11)
	c) M2 DNA polymerase*	(12)
	d) T4 DNA polymerase	(13)
	3. Eukaryotic DNA polymerases	
	a) Human DNA polymerase alpha	(14)
	b) Yeast DNA polymerase I	(15)
	c) Yeast DNA polymerase II	(16)
	d) Yeast DNA polymerase III (delta)	(17)
	e) Yeast DNA polymerase Rev3	(18)
	4. Viral DNA polymerases	
	a) Herpes-1 DNA polymerase	(19)
	b) Human cytomegalovirus DNA polymerase	(20)
	c) Epstein-Barr virus DNA polymerase	(21)
	d) Varicella-Zoster virus DNA polymerase	(22)
	e) Fowlpox virus DNA polymerase	(23)
	f) Vaccinia virus DNA polymerase	(24)
	g) Autographa californica nuclear polyhedrosis virus (AcMNPV) DNA polymerase	(25)
	h) Adenovirus-2 DNA polymerase*	(26)
	i) Adenovirus-7 DNA polymerase*	(27)
	j) Adenovirus-12 DNA polymerase*	(28)
	5. Eukaryotic linear DNA plasmid encoded DNA polymerases	
	a) S-1 maize mitochondrial DNA polymerase*	(29)
	b) *Kluyveromyces lactis* plasmid pGKL1 DNA polymerase*	(30)
	c) *Kluyveromyces lactis* plasmid pGKL2 DNA polymerase*	(31)
	d) *Claviceps purpurea* plasmid pCLK1 DNA polymerase*	(32)
	e) *Ascobolus immersus* plasmid pAI2 DNA polymerase*	(33)
C.	**Family C DNA polymerases**	
	Bacterial replicative DNA polymerases	
	a) *E. coli* DNA polymerase III α subunit	(34)
	b) *Salmonella typhimurium* DNA polymerase III α subunit	(35)
	c) *Bacillus subtilis* DNA polymerase III	(36)
D.	**Family X DNA polymerases**	
	a) Rat DNA polymerase β	(37)
	b) Human DNA polymerase β	(38,39)
	c) Human terminal deoxynucleotidyltransferase (TdT)	(40)
	d) Bovine terminal deoxynucleotidyltransferase (TdT)	(41)
	e) Mouse terminal deoxynucleotidyltransferase (TdT)	(41)

Table 1. The main families and subclassifications of DNA polymerases. Those DNA polymerases marked with a star (*) are protein-primed DNA polymerases.

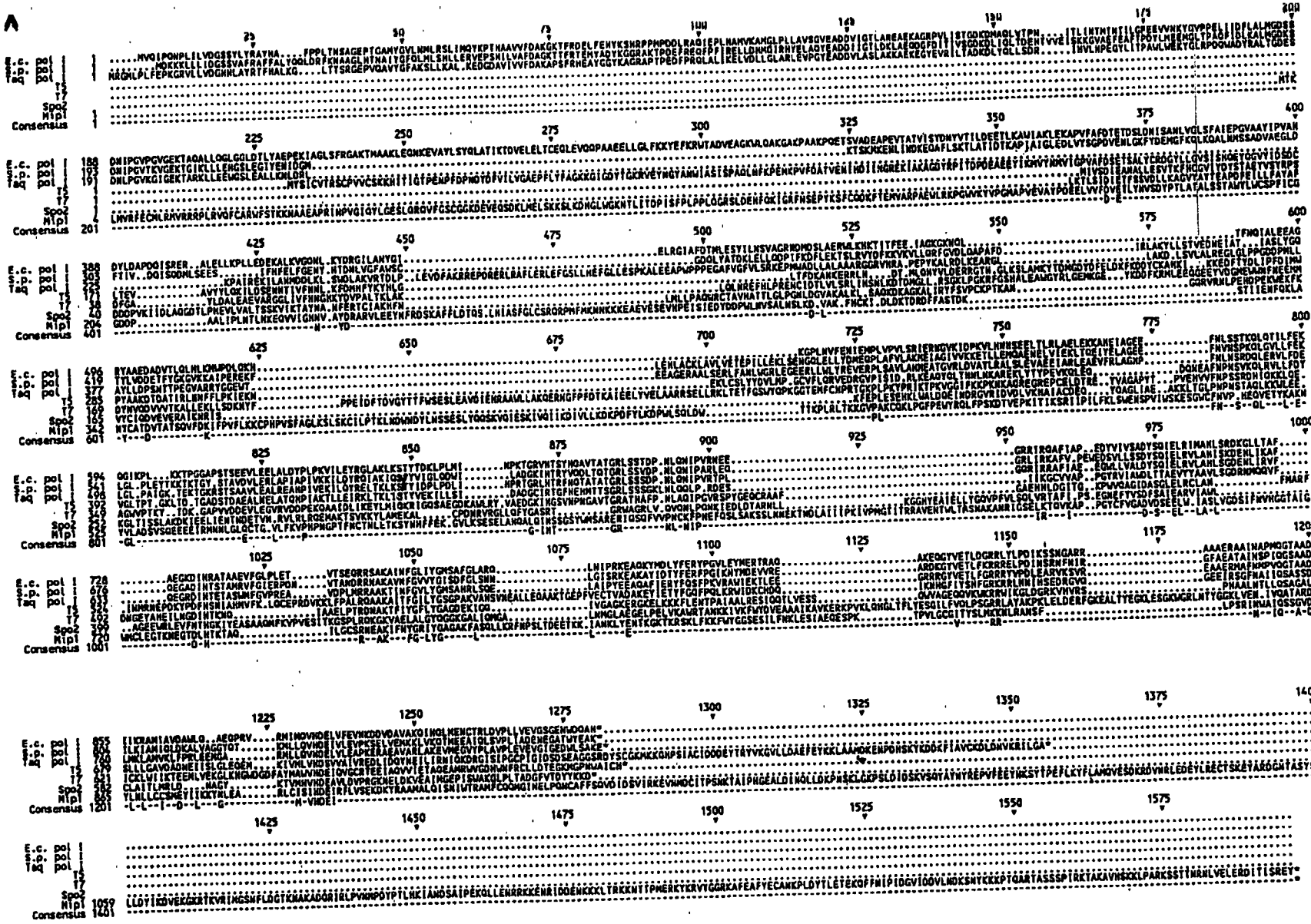
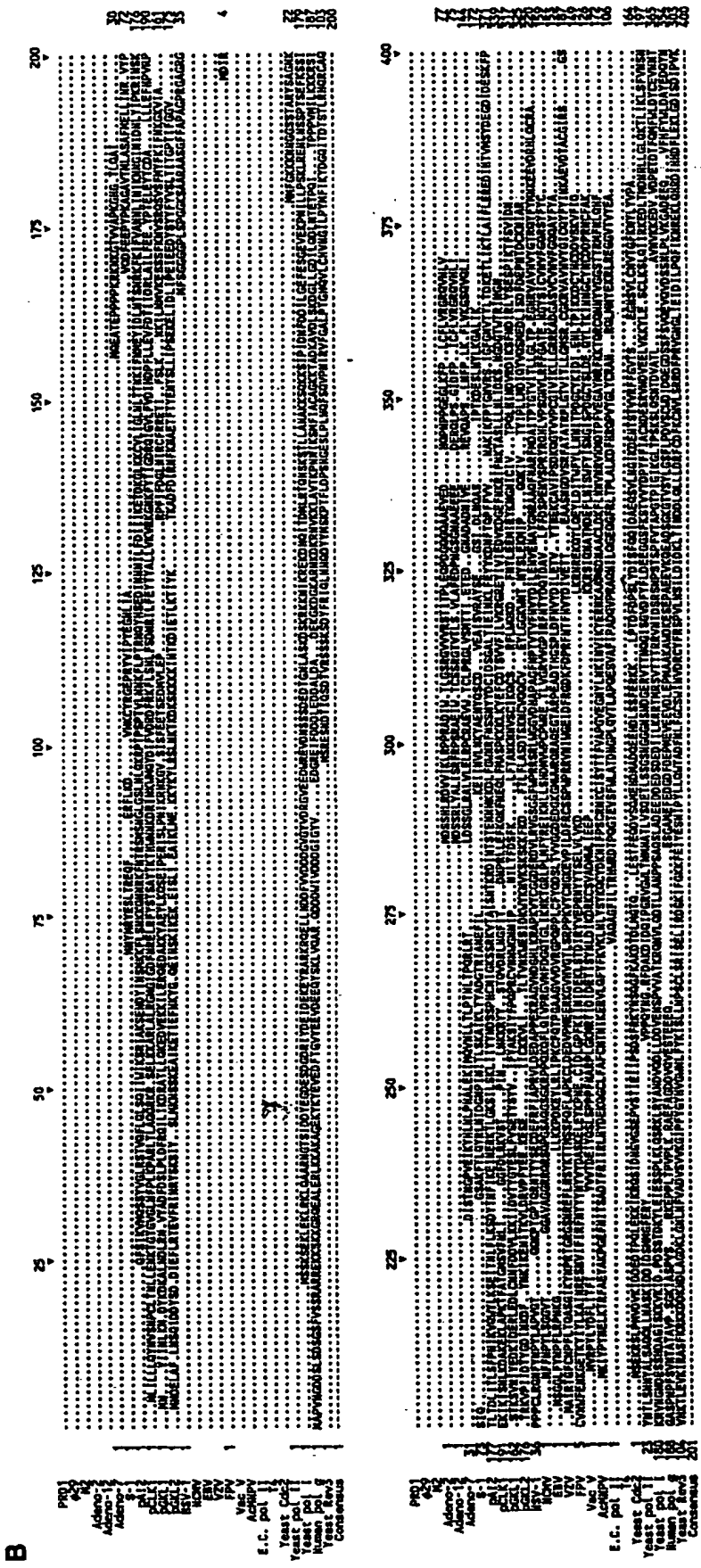
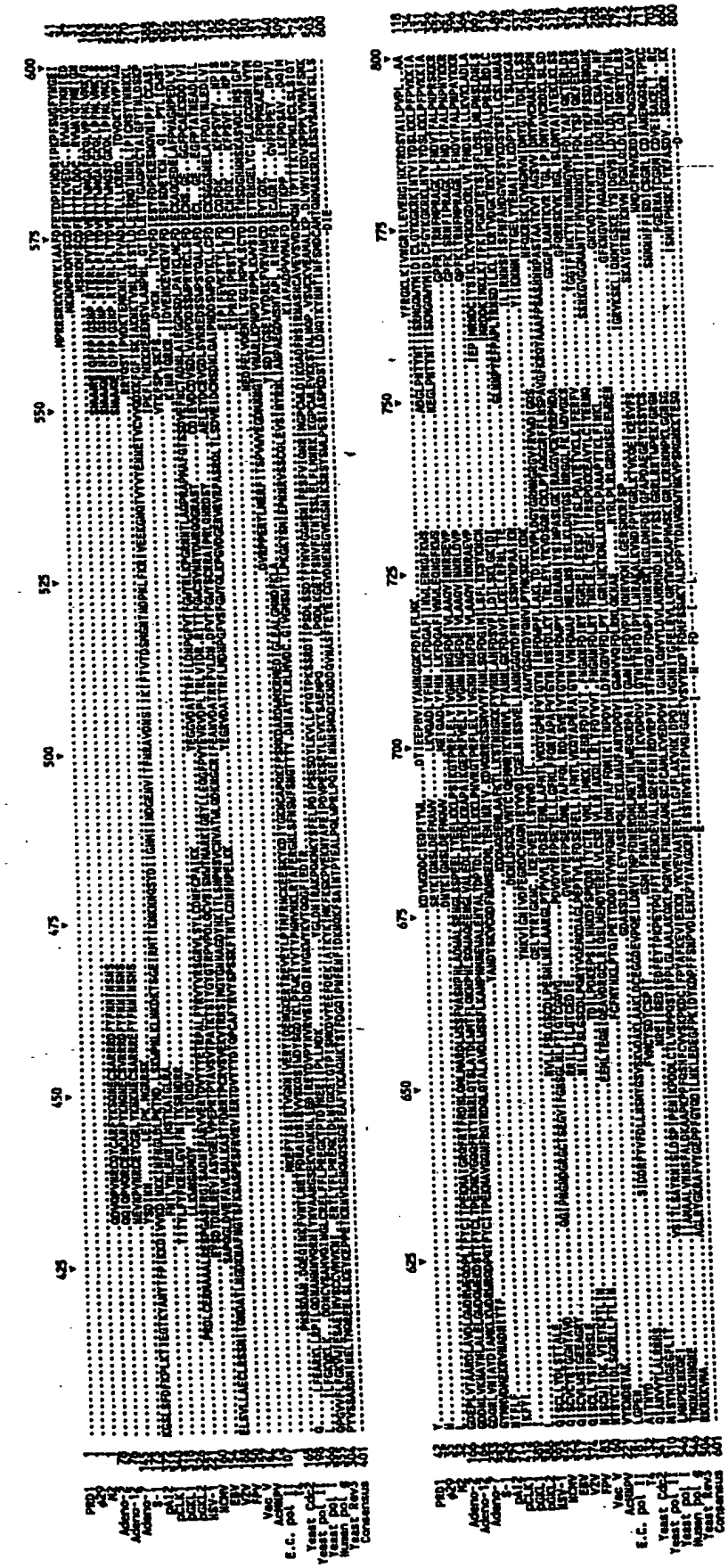
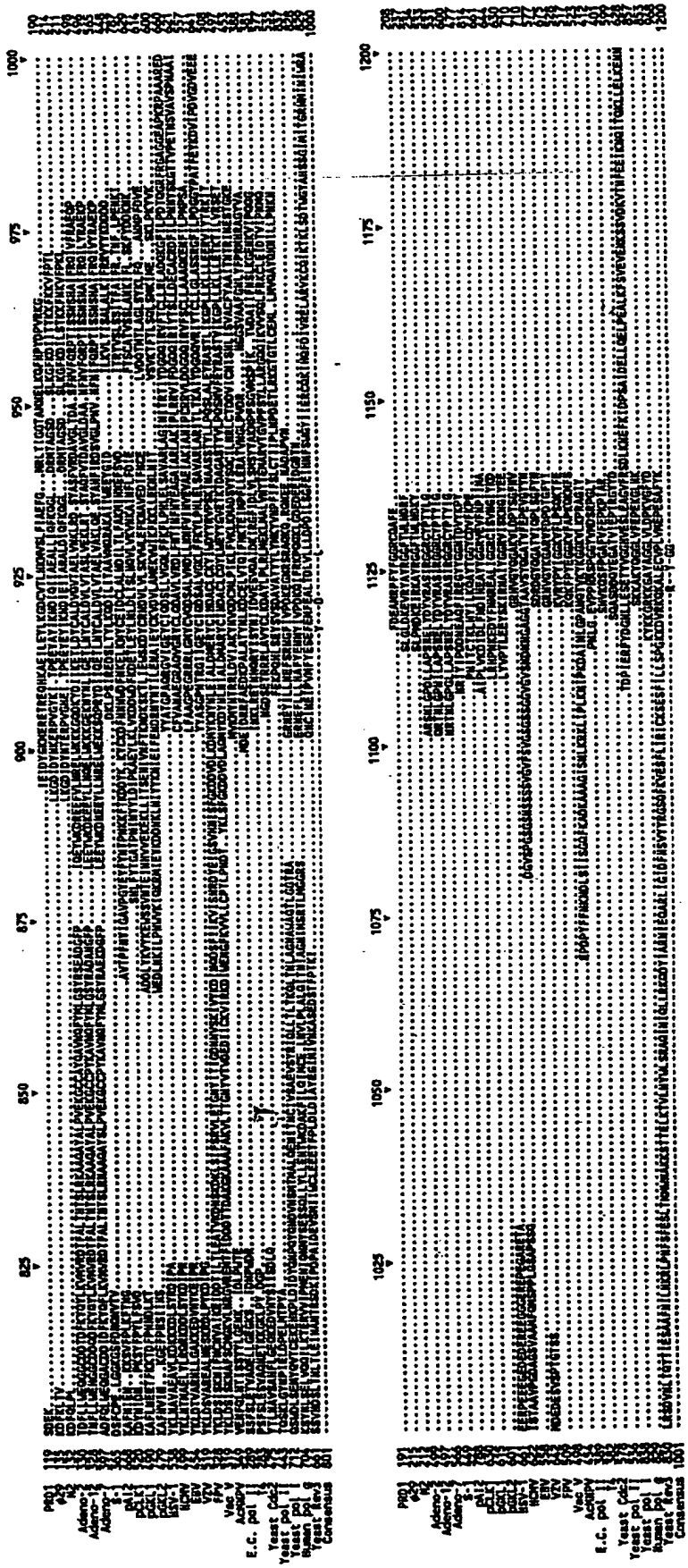
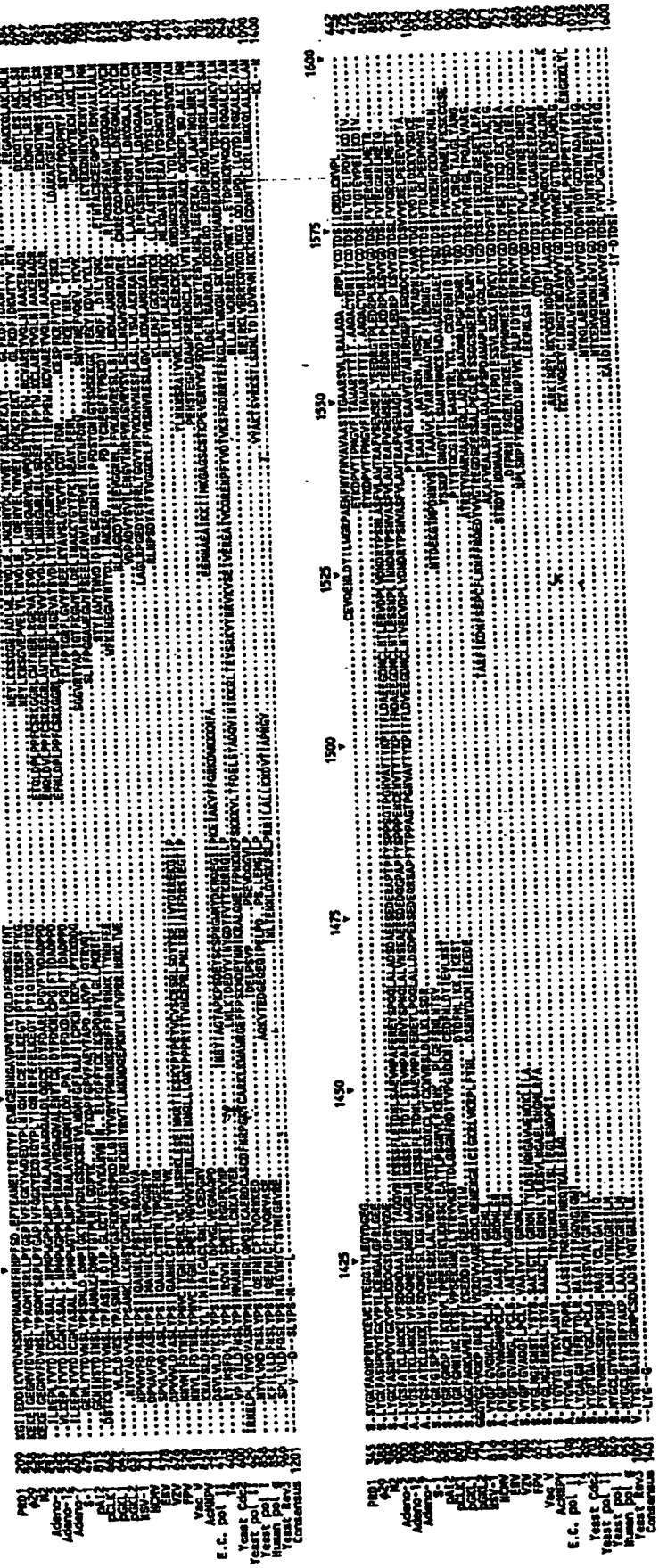


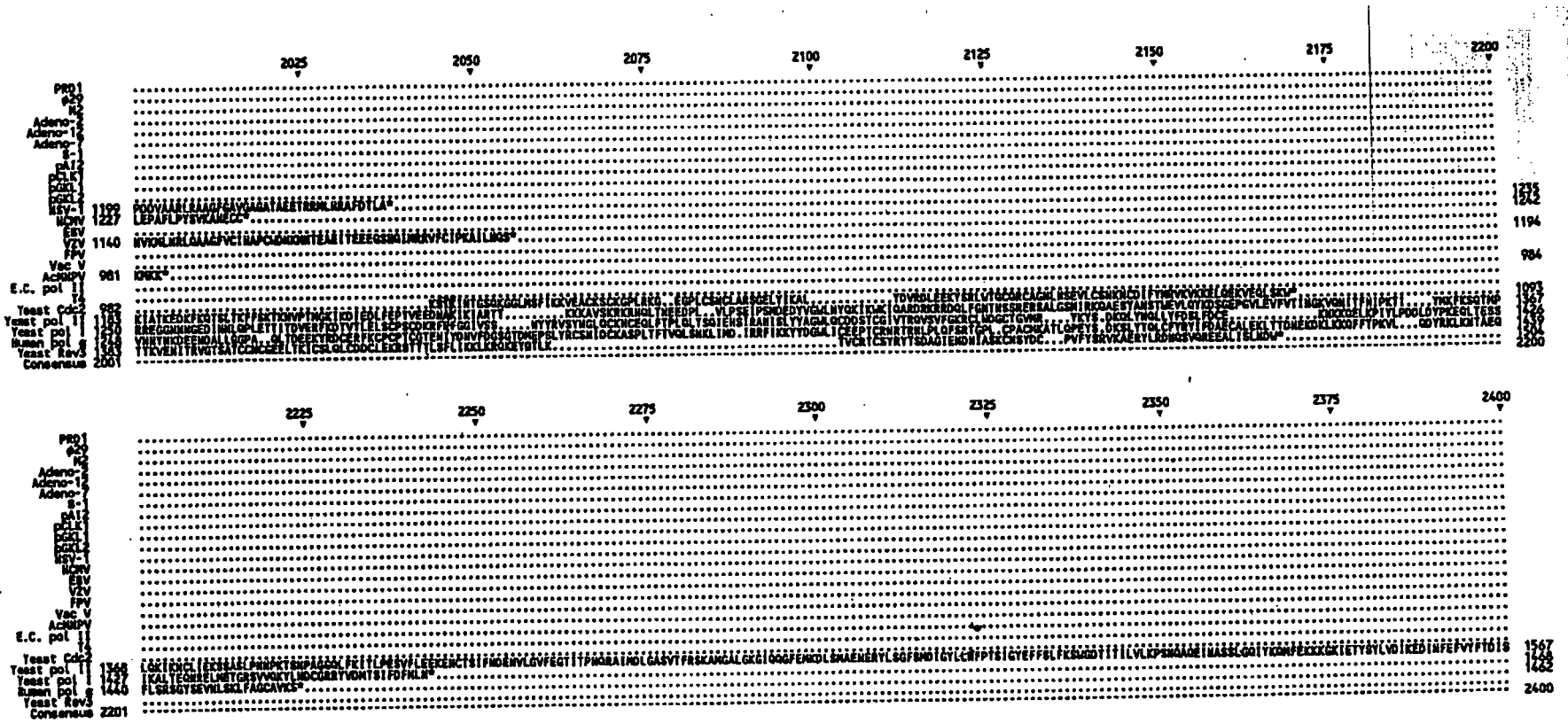
Figure 1A. Family A DNA polymerases—*E. coli* DNA pol. I (E.c. pol I)(1), *Streptococcus pneumoniae* DNA pol. I (S.p. pol I)(2), *Thermus aquaticus* DNA pol. I (Taq pol I)(3), T5 DNA pol. (4), T7 DNA pol. (5), Spo2 DNA pol. (6), and yeast mitochondrial DNA pol. (MIP1)(7).

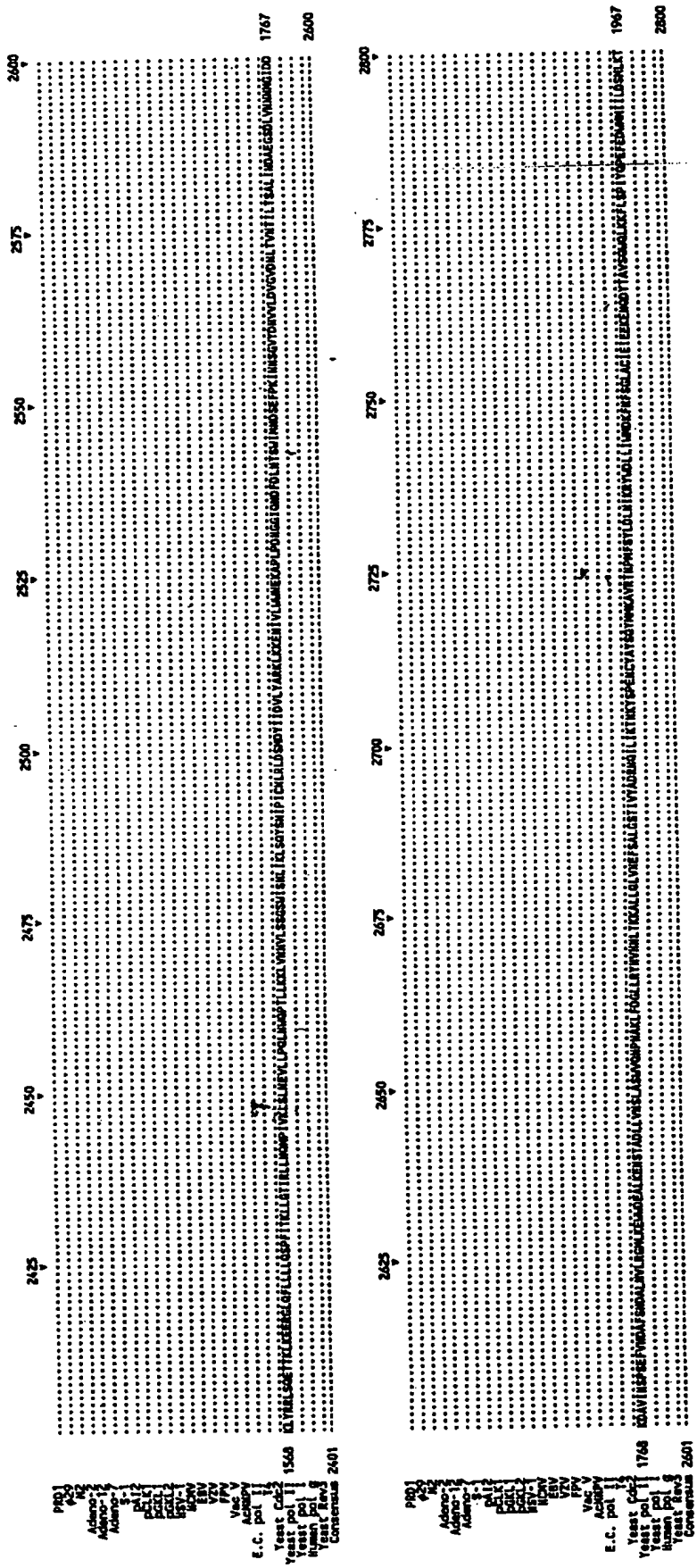

B
PRD1
M2
T4
Adeno-1
Adeno-2
S-
PolI
PolII
PolIII
RSV-1
HCMV
EBV
VZV
FPV
Vac V
AcNPV
E.c. pol I
Yeast Cdc2
Yeast pol II
Human pol II
Yeast Rev3
Consensus








2025
2050
2075
2100
2125
2150
2175
2200
2225
2250
2275
2300
2325
2350
2375
2400
PRD1
φ29
M2
Adeno-2
Adeno-12
Adeno-7
β-
pAI2
PLL1
PGL1
PGL2
REV-1
HCMV
EBV
VZV
FPV
Vac V
AcNMPV
E.C. pol II
Yeast Cdc2
Yeast pol II
Yeast pol I
Human pol α
Yeast RevS
Consensus



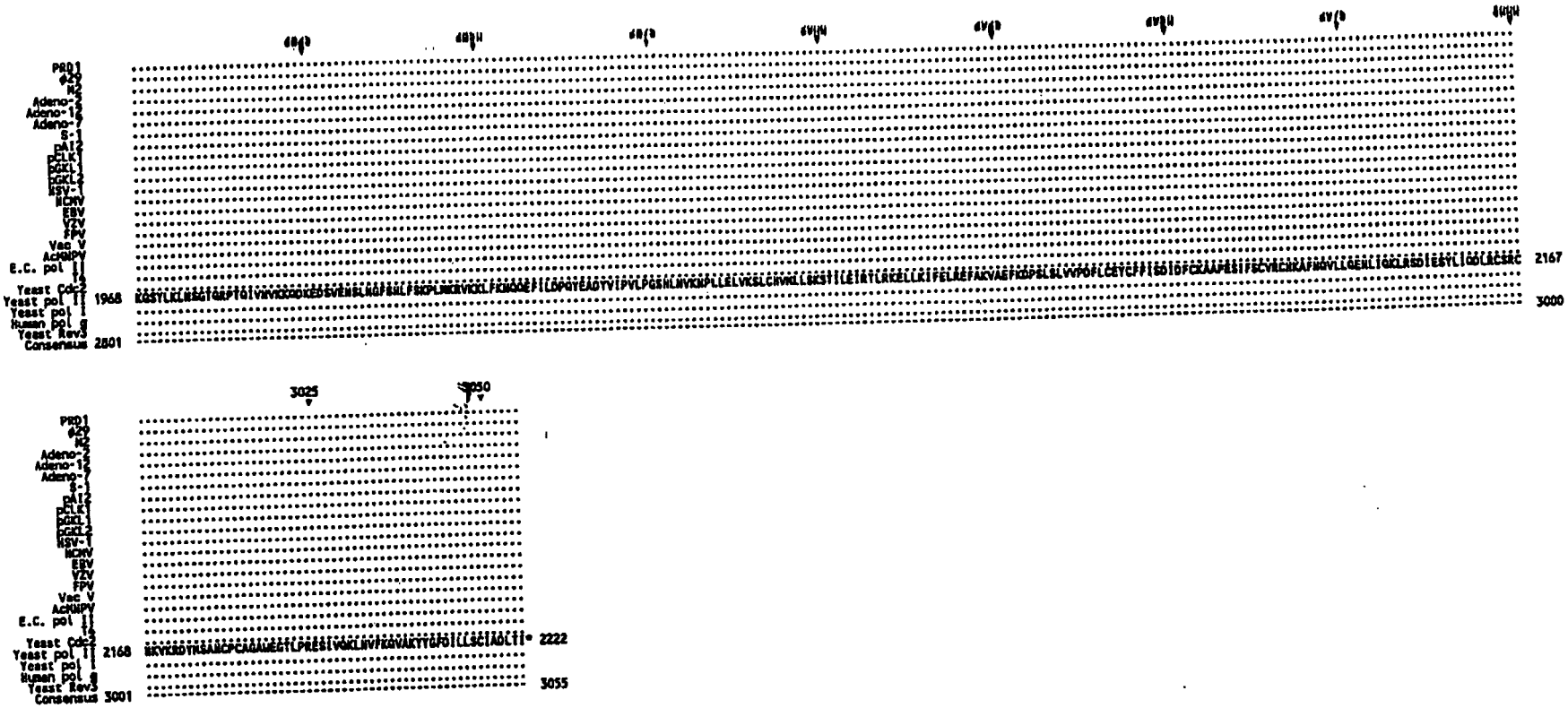


Figure 1B. Family B DNA polymerases—PRD1 DNA pol. (9,10), ϕ29 DNA pol. (11), M2 DNA pol. (12), Adenovirus type-2 DNA pol. (Adeno-2)(26), Adenovirus type-12 DNA pol. (Adeno-12)(28), Adenovirus type-7 DNA pol. (Adeno-7)(27), S-1 maize mitochondrial DNA pol. (S1)(29), *Ascobolus immersus* plasmid DNA pol. (pAI2)(33), *Claviceps purpurea* plasmid DNA pol. (pCLK1)(32), *Kluyveromyces lactis* plasmid DNA pol. (pGKL1)(30), *kluyveromyces lactis* plasmid DNA pol. (pGKL2)(31), herpes simplex type-1 DNA pol. (HSV-1)(19), Human cytomegalovirus DNA pol. (HCMV)(20), Epstein-Barr virus DNA pol. (EBV)(21), Varicella-zoster virus DNA pol. (VZV)(22), Fowlpox virus DNA pol. (FPV)(23), Vaccinia virus DNA pol. (VacV)(24), Autographa californica nuclear polyhedrosis virus DNA pol. (AcMNPV)(25), *E. coli* DNA pol. II (E.c. pol II)(8), T4 DNA pol. (13), Yeast DNA pol. III (CDC2)(17), Yeast DNA pol. II (16), Yeast DNA pol. I (15), Human DNA pol. α (14), and Yeast Rev3 DNA pol.(18).

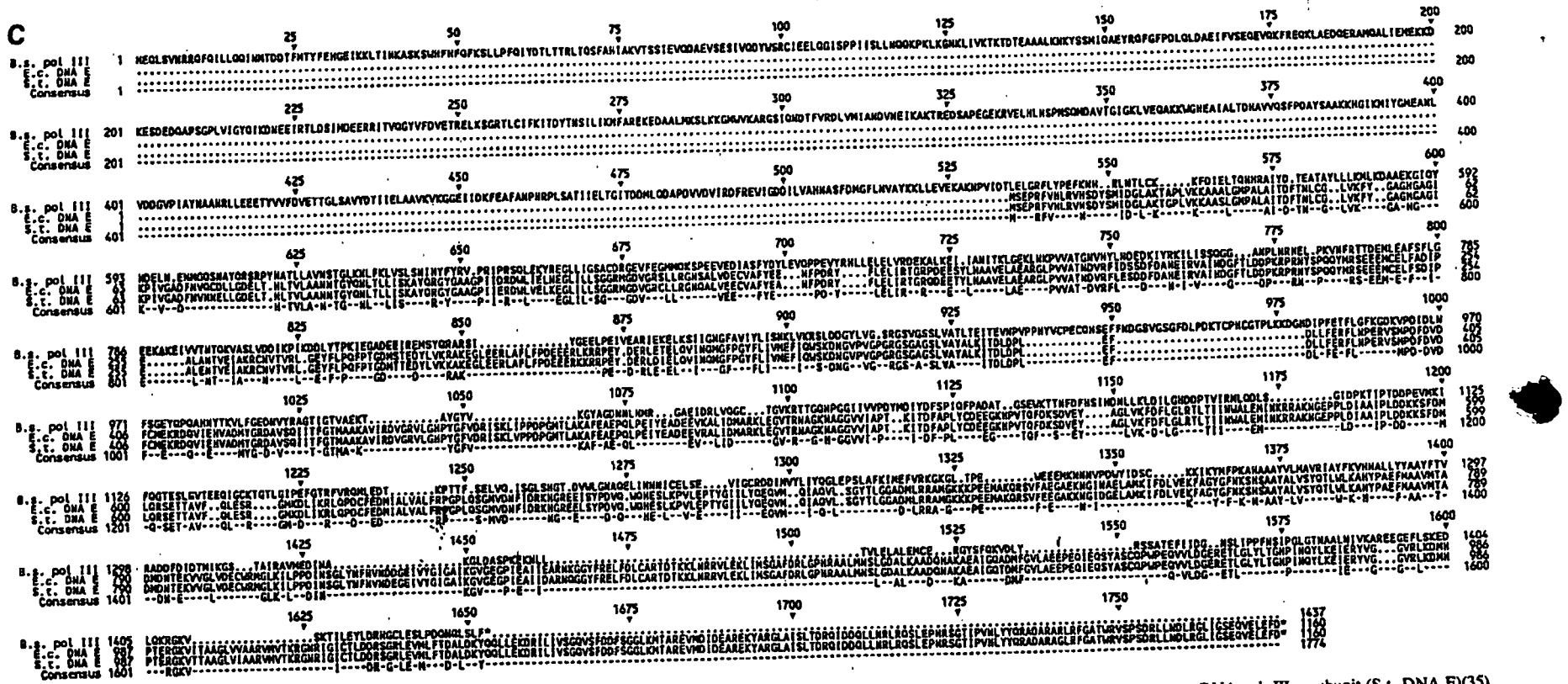


Figure 1C. Family C DNA polymerases—*Bacillus subtilis* DNA pol. III (B.s. pol III)(36), *E. coli* DNA pol. III α-subunit (E.c. DNA E)(34), and *Salmonella typhimurium* DNA pol. III α-subunit (S.t. DNA E)(35).

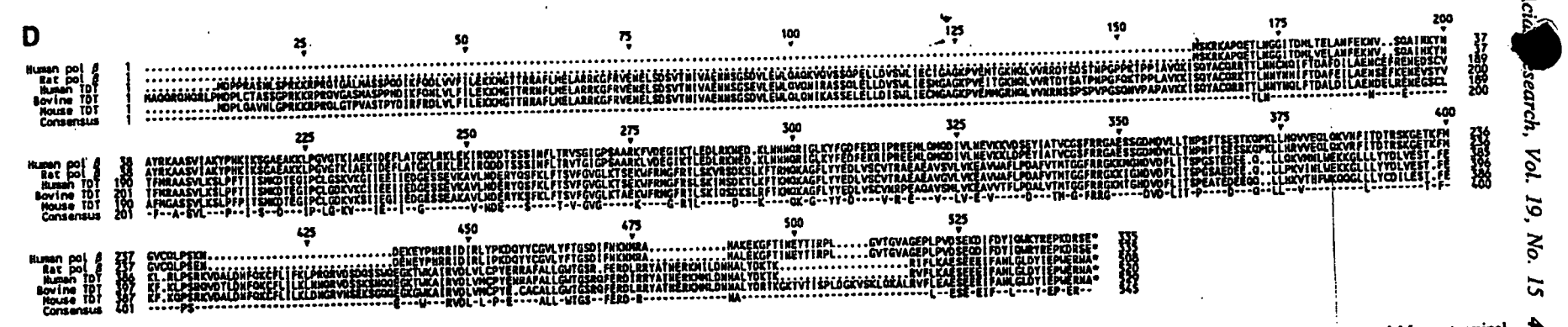


Figure 1D. Family X DNA polymerases—Human DNA pol. β (38,39), Rat DNA pol. β (37), Human terminal transferase (Human TDT)(40), Bovine terminal transferase (Bovine TDT)(41), and Mouse terminal transferase (Mouse TDT)(41).

The EMBO Journal vol.7 no.1 pp.37 - 47, 1988

EXHIBIT C

Human DNA polymerase α gene expr ssion is c ll proliferation dependent and its primary structure is similar to both prokaryotic and eukaryotic replicative DNA polymerases

Scott W.Wong, Alan F.Wahl, Pau-Miau Yuan[1], Naoko Arai[2], Barbara E.Pearson, Ken-ichi Arai[2], David Korn, Michael W.Hunkapiller[1] and Teresa S.-F.Wang

Laboratory of Experimental Oncology, Department of Pathology, Stanford Medical School, Stanford University, Stanford, CA 94305, [1]Applied Biosystems, Foster City, CA 94404 and [2]DNAX Research Institute of Molecular and Cellular Biology, Palo Alto, CA 94304, USA

Communicated by A.Kornberg

We have isolated cDNA clones encoding the human DNA polymerase α catalytic polypeptide. Studies of the human DNA polymerase α steady-state mRNA levels in quiescent cells stimulated to proliferate, or normal cells compared to transformed cells, demonstrate that the polymerase α mRNA, like its enzymatic activity and *de novo* protein synthesis, positively correlates with cell proliferation and transformation. Analysis of the deduced 1462-amino-acid sequence reveals six regions of striking similarity to yeast DNA polymerase I and DNA polymerases of bacteriophages T_4 and $\phi 29$, herpes family viruses, vaccinia virus and adenovirus. Three of these conserved regions appear to comprise the functional active site required for deoxynucleotide interaction. Two putative DNA interacting domains are also identified.

Key words: primary structure/replicative DNA-polymerases/sequence similarity/structural gene/transcription

Introduction

Cell proliferation and the transmission and maintenance of error-free genetic information from one generation to the next are dependent on the mechanism of DNA replication. Genomic DNA replication is a complex and tightly regulated process involving the orderly coordination of many protein−protein and protein−DNA interactions (Kornberg, 1980, 1982). A key component of the chromosomal replication apparatus is DNA polymerase α, which is generally accepted as the principal polymerase involved in eukaryotic DNA replication (Kornberg, 1980, 1982; Campbell, 1986; Fry and Loeb, 1986). Many lines of evidence support this concept: its enzymatic activity positively correlates with DNA synthesis during cell proliferation (Fry and Loeb, 1986); all its specific inhibitors also inhibit DNA replication *in vivo* (Ikegami *et al.*, 1978; Fry and Loeb, 1986); a mutant which is temperature sensitive for DNA synthesis has been identified as a DNA polymerase α mutant (Murakami *et al.*, 1985); and monoclonal antibodies specific for DNA polymerase α inhibit DNA synthesis in permeabilized cells or when microinjected into nuclei (Miller,M.R. *et al.* 1985; Miller *et al.*, 1986). The recent reports that DNA polymerase α plays a central role in SV40 DNA replication *in vitro* (Li and Kelly, 1984, 1985; Stillman and Gluzman, 1985; Wobbe *et al.*, 1985; Murakami *et al.*, 1986), a model system of eukaryotic DNA replication, further underscore the importance of this enzyme.

DNA polymerase α lacks the 3′−5′ proof reading exonuclease which is required to remove mismatched nucleotides during DNA polymerization (Kornberg, 1980, 1982; Fry and Loeb, 1986). However, another eukaryotic DNA polymerase, designated δ, has been identified (Brynes *et al.*, 1976), which possesses a 3′−5′ exonuclease activity and, like polymerase α, is sensitive to the inhibitor aphidicolin (Lee *et al.*, 1984). The relationship between polymerases α and δ is unknown. The identification of a cellular protein, proliferating cell nuclear antigen (PCNA), required for efficient viral DNA chain-elongation *in vitro*, and the discovery that this protein is able to stimulate DNA polymerase δ but not polymerase α activity, raises the question of whether there are two polymerases involved in eukaryotic DNA replication (Bravo *et al.*, 1987; Perlich *et al.*, 1987a,b). The recent finding of a cryptic proof reading 3′−5′ exonuclease activity associated with the catalytic polypeptide of *Drosophila* embryo DNA polymerase α, when separated from other subunits (Cotterill *et al.*, 1987), further stimulates interesting questions about the relationship between these two DNA polymerases.

Despite more than a quarter century of biochemical characterization of DNA polymerase α (Fry and Loeb, 1986) little is known about the regulation of the expression of this essential DNA replication enzyme, which nucleotide structural elements may be responsible for the cell-proliferation-dependent expression and whether its expression is the direct target of cascading biochemical events induced by growth factors or mitogens. In addition, nothing is known about the structure−function relationships of DNA polymerase α protein domains required for substrate recognition, or for the orderly coordination of protein−protein and protein−DNA interactions during chromosome replication. In an attempt to address these questions and to define the relationships between DNA polymerase α and δ, a near full-length cDNA of the human DNA polymerase α catalytic polypeptide has been isolated.

Comparison of the steady-state mRNA levels of quiescent cells stimulated to proliferate, or normal human cells compared to transformed cells, demonstrates that the previously reported increase of enzymatic activity during cell proliferation or transformation correlates with the level of steady-state mRNA. Analysis of the deduced primary structure of human DNA polymerase α with several viral DNA polymerases, yeast DNA polymerase I, *Escherichia coli* bacteriophage T_4 and *Bacillus* phage $\phi 29$ DNA polymerase identifies six regions highly conserved among these polymerases. Three of these conserved domains appear to comprise th functionally active sites required for deoxynucleotide interaction and two other regions are postulated

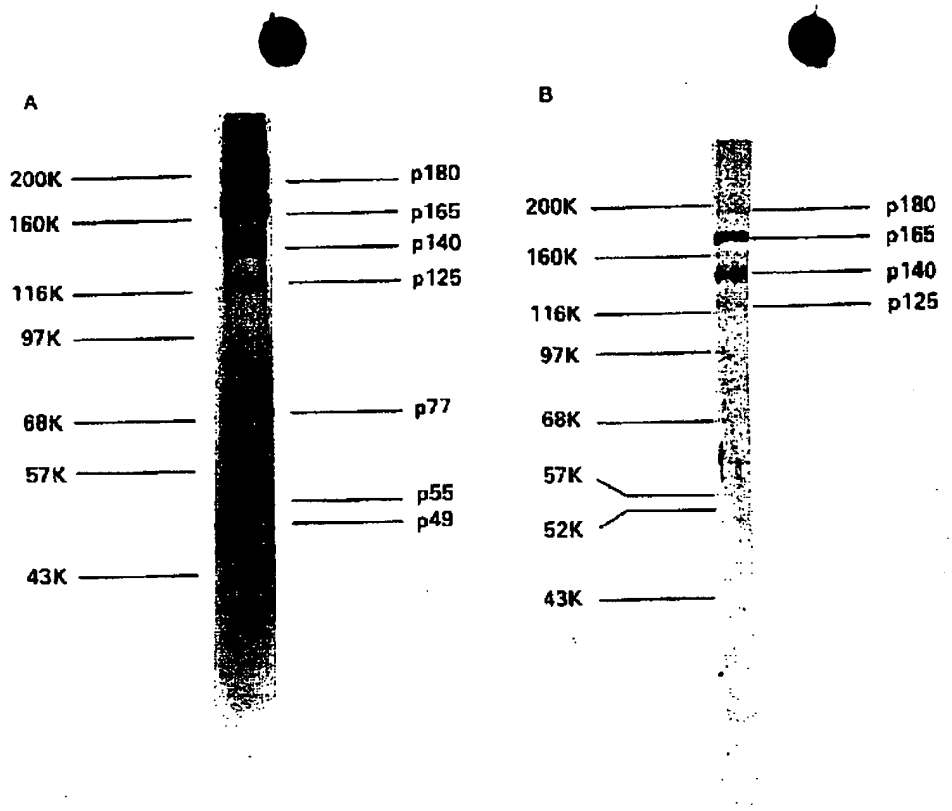


Fig. 1. SDS gel analysis of human DNA polymerase α polypeptides. (A) Immunopurified human DNA polymerase α. (B) Isolated human DNA polymerase α catalytic polypeptides. 1% of the sample before and after HPLC separation as described under Materials and methods was analyzed on an SDS–8% polyacrylamide slab gel and stained with silver.

Table I. Human DNA polymerase α peptide sequence analysis

Cycle no.	T9		T19		T23		T24		T25		T264		T265	
	Res	pmol	Res	pmol	Res	pmol	Res	pmol	Res	pmol	Res	pmol	Res	pmol
1	S	20	A	92	G	24	N	17	Y	50	D	10.1	Q	3.9
2	G	54	A	96	P	26	Y	23	I	53	T	1.7	D	0.7
3	Y	58	Y	62	C	5	A	25	F	51	G	7.8	N	2.5
4	S	11	A	76	W	8	F	24	D	22	N	4.4	L	5.3
5	E	41	G	40	L	25	E	12	A	48	F	8.5	T	0.7
6	V	46	G	37	E	14	I	24	E	37	V	7.2	I	4.3
7	N	24	L	56	V	19	P	14	C	10	I	7.9	D	3.8
8	L	30	V	56	K	5	D	8	A	48	G	7.4	T	0.7
9	S	8	L	44			V	24	L	45	Q	3.4	Q	2.7
10	K	15	D	18			P	10	E	22	I	6.5	Y	1.6
11			P	20			E	5	K	17	L	5.5	Y	2.2
12			K	19			K	8			S	0.6	L	4.2
13											D	2.3	A	2.6
14											Q	1.4	Q	3.6
15											S	0.4	Q	3.1
16											R	<<	I	2.4

Shown are the PTH-amino acids observed at each sequence cycle. << indicates yield too low for quantitation.

to be DNA binding domains. the presence of these conserved amino acid sequences among replicative DNA polymerases from phylogenetically distant species suggests they all may have evolved from a single primordial gene.

Results

Purification of the catalytic polypeptide and protein sequencing

The advancement of utilizing monoclonal antibodies specifically against human DNA polymerase α for immunoaffinity purification has defined the protein structure and subunit components of this enzyme (Wang *et al.*, 1984; Wong *et al.*, 1986). DNA polymerase α contains (i) the catalytic polypeptide which, *in vitro*, is a family of large phosphopolypeptides of 180–125 kd, previously demonstrated by tryptic peptide mapping to be derivatives of the same primary structure (Wong *et al.*, 1986); (ii) a 77 kd phosphoprotein of unknown function; and (iii) two polypeptides of 55 and 49 kd reported to be associated with DNA primase activity (Tseng and Ahlem, 1983) (Figure 1A). The catalytic polypeptides of human polymerase α were separated from associated

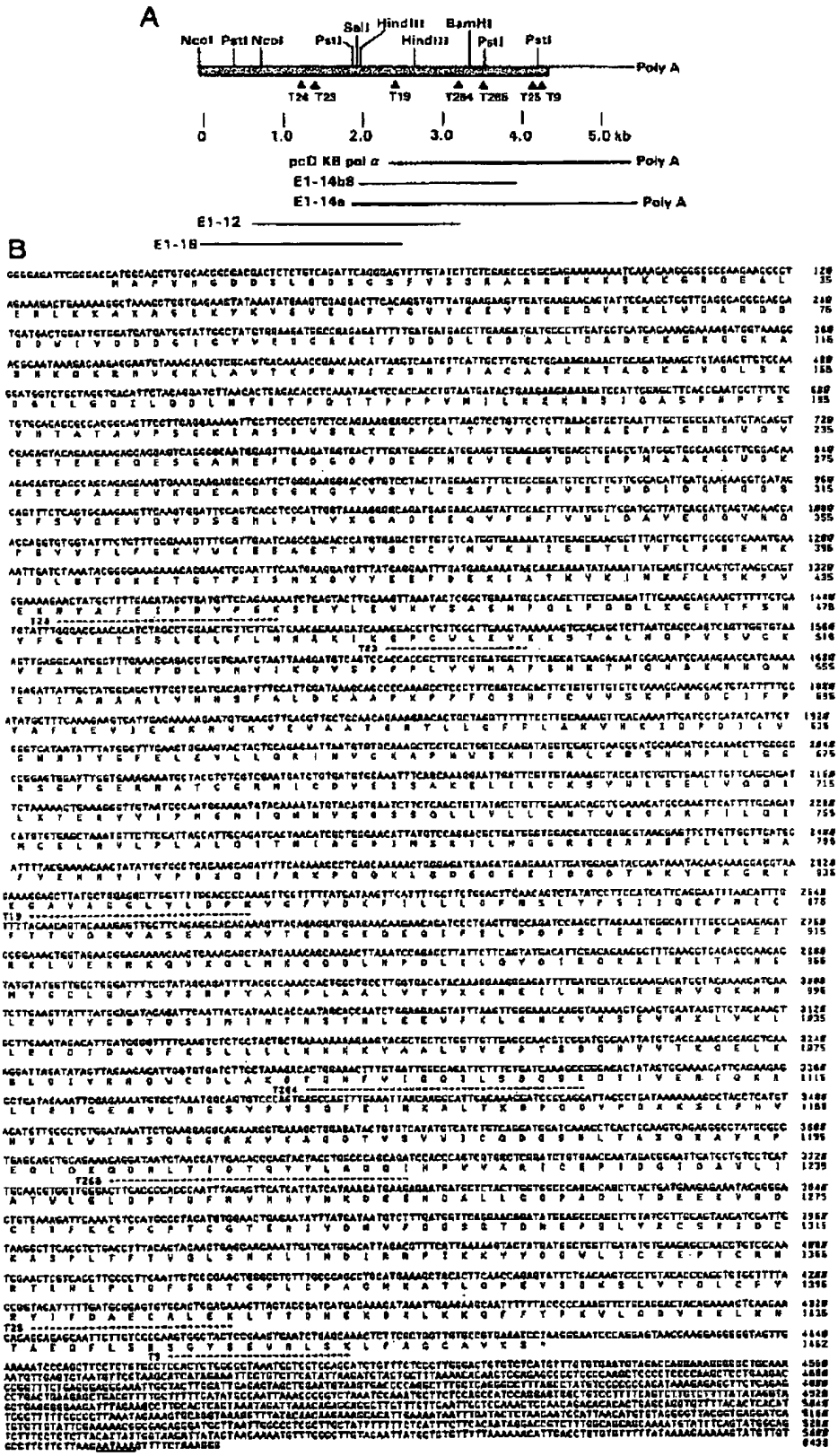


Fig. 2. Human DNA polymerase α cDNA. (A) Restriction map of human DNA polymerase α and overlapping cDNA clones. The stippled box represents the coding region of human DNA polymerase α and the solid line indicates the 5' and 3' non-coding region. ▲ indicates the locations of each of the previously determined amino acid sequences. The five overlapping cDNA clones are pcD-KBpolα, E1-14b8, E1-14a, E1-12 and E1-19. (B) Nucleotide sequence and deduced amino acid sequence of human DNA polymerase α. Nucleotides are numbered at the upper right and amino acids at the lower right. Peptide sequences derived from immunopurified human DNA polymerase α preparation and used to design oligonucleotide probes as described under Materials and methods are underlined with dotted lines and labeled according to Table I. Amino acid number starts at methionine 1. * indicates the termination codon TAA.

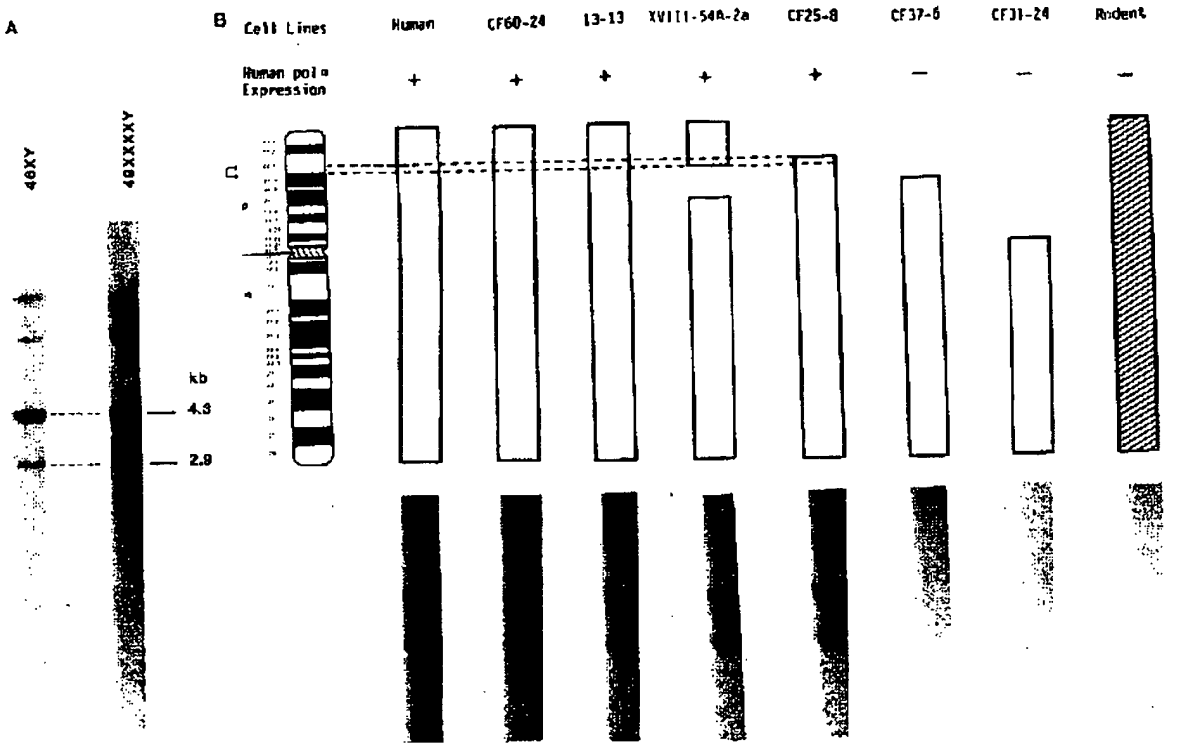


Fig. 3. Physical mapping of human DNA polymerase α cDNA on X chromosome. (A) Genomic Southern blot. 10 μg of human genomic DNA isolated from 46XY (1X normal male) and 49XXXXY (4X, from cell line GM1202A) were digested with *Eco*RI and hybridized as described under Materials and methods. (B) Human DNA polymerase α structural gene mapping. The idiogram of trypsin–Giemsa banding pattern of human X chromosome is presented at the left. *Eco*RI digested genomic DNA, 10 μg each from human KB cells, rodent AKR thymoma cell line BW5147 and somatic hybrids (Wang *et al.*, 1985): CF60-24 containing human chromosome 4, 13, 17, 19, 21 and X; hybrid 13–13 containing only human X chromosome; hybrid clone XVIII-54A-2a containing *de novo* interstitial deletion of most of the Xp21 band [46X del(X) (pter--21.3 :: p21.1–qter)]; and the other three CF series hybrid clones containing X/autosome translocations: CF25-8 containing an X/13 translocation, retained der(X), Xqter--Xp22.1; CF37-6 containing an X/11 translocation, retained der(X), Xqter--Xp21.2; CF31-24 containing an X/20 translocation with dr(X), Xqter--Xcent, were hybridized at 42°C with 30 ng ^{32}P-labeled (10^9 c.p.m./μg) *Pst*I 700-bp fragment of pcD-KB*polα*. Hybridization and washing conditions were as described under Materials and methods.

subunits by gel permeation HPLC columns (Figure 1B). The separated catalytic polypeptides were pooled and treated as a single entity, digested with trypsin and fractionated by preparative reverse-phase HPLC. The amino acid sequences of seven peptides were determined (Table I) as described in Materials and methods. In all, the sequences of 85 amino acids were established and used to design single, long antisense oligonucleotide probes (Lathe, 1985) by which a near full-length cDNA clone of human DNA polymerase α was isolated.

Primary structure of human DNA polymerase α

The 5433 nucleotides of the human polymerase α cDNA contain a single open reading frame coding for 1462 amino acids (Figure 2A and B). An in-frame initiator ATG codon flanked by nucleotides matching Kozak's criteria for a translation initiation site was identified (Figure 2B) (Kozak, 1981). All seven experimentally determined human DNA polymerase α peptide sequences listed in Table I are identified within this amino acid sequence (Table I, Figure 2A and B). Based on the deduced amino acid sequence, the estimated M_r of the recombinant polymerase α is 165 kd. Primer extension with two synth tic oligonucleotides corresponding to two separate regions of the 5′-end localizes the transcription start site 295 nucleotides upstream from the putative translation initiation codon (data not shown).

Localization of the structural gene

The human DNA polymerase α gene was previously mapped by expression to a single genetic locus on the short arm of the X chromosome at the junctional region of Xp21.3 to Xp22.1 (Wang *et al.*, 1985). The cDNA insert was shown to be X chromosome-linked by comparative genomic Southern hybridization with normal male DNA (46 XY) and DNA from a cell line of 4X (karyotyped 49 XXXXY) DNA (Figure 3A). Two *Eco*RI-digested genomic DNA bands are observed by hybridization with a *Pst*I restriction fragment of pcD-KB*polα* (Figure 2A), both resulting in 1:4 ratio of signal intensity (Figure 3A). Using this *Pst*I restriction fragment of pcD-KB*polα* cDNA clone, the chromosomal localization of the DNA polymerase α structural gene was analyzed directly by Southern hybridization of *Eco*RI-digested genomic DNA samples from a panel of human–rodent hybrids containing either an intact human X chromosome, different but overlapping regions of the human X chromosome and a hybrid clone with an interstitial deletion of the human X chromosome (Wang *et al.*, 1985). Under conditions that exclude cross-hybridization to rodent

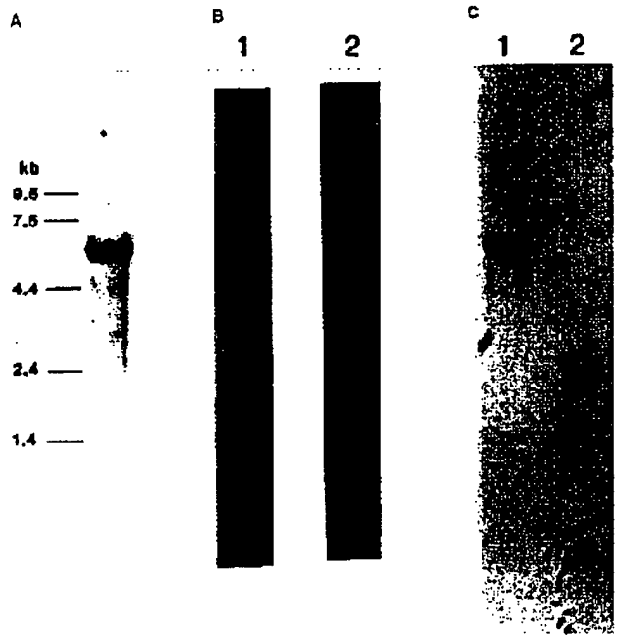


Fig. 4. Steady-state mRNA analyses from quiescent cells stimulated to proliferate and comparison of normal and transformed cells. (A) Northern hybridization analysis of human DNA polymerase α mRNA. 10 μg of polyadenylated mRNA from early mid-log human KB cells was hybridized with 50 ng of ^{32}P-labeled *Hin*dIII/*Bam*HI 700-bp restriction fragment as described in Materials and methods. RNA sizes are given in kb and were determined by staining of parallel blot with RNA standard ladder marker from Bethesda Research Laboratory (BRL). (B) Steady-state mRNA from quiescent and proliferative cells. 25 μg of total RNA isolated from normal human lung fibroblast (IMR90), cultured in the 0.1% fetal calf serum (FCS) for 96 h to effect quiescence (lane 1) or stimulated to proliferate by activation with 10% FCS for 30 h (lane 2) were hybridized with 100 ng of ^{32}P-labeled anti-sense DNA polymerase α ribo-probe (5×10^8 c.p.m./μg) as described in Materials and methods. Autoradiography was for 5 days with intensifier screen at −70°C. (C) Steady-state mRNA comparison of transformed and normal human cells. 5 μg of poly(A)$^+$ mRNA from transformed cell line, MDA4, (lane 1) and from normal human proliferating tissue isolated from 20 week placenta (lane 2) were compared for relative abundance of steady state DNA polymerase α mRNA as described in Materials and methods.

Table II. DNA polymerase α activity in transformed and non-transformed human cells

Transformed cells	Units per 10^7 cells	Non-transformed cells	Units per 10^7 cells
293	19.0	TNHF	2.1
KB	14.0	GM1604	1.7
		IMR90 (midlog)	2.4
		IMR90 (quiescent)	UD

Units of DNA polymerase are defined as nmol of labeled dAMP incorporated/h at 37°C in the presence and absence of 5 μg/ml aphidicolin. The results are averages of duplicate determinations. UD represents undetectable background value.

DNA, the structural gene for human DNA polymerase α was mapped precisely to the previously determined expression locus at Xp21.3 to Xp22.1 (Figure 3B).

Analysis of steady-state mRNA during cell proliferation and transformation

Characterization of this cDNA insert by Northern hybridization yields a single 5.8 kb band (Figure 4A) which is sufficient to encode a polypeptide of 165−180 kd. Northern hybridization of restriction fragments from each of the overlapping cDNA clones or the 5′-end of the near full-length cDNA clone all result in a single mRNA hybridization signal of 5.8 kb. These results indicate there is neither usage of multiple polyadenylation addition sites, generating mRNA of variable length from a single gene, nor various splicing events or processing occurring to generate multiple mRNAs.

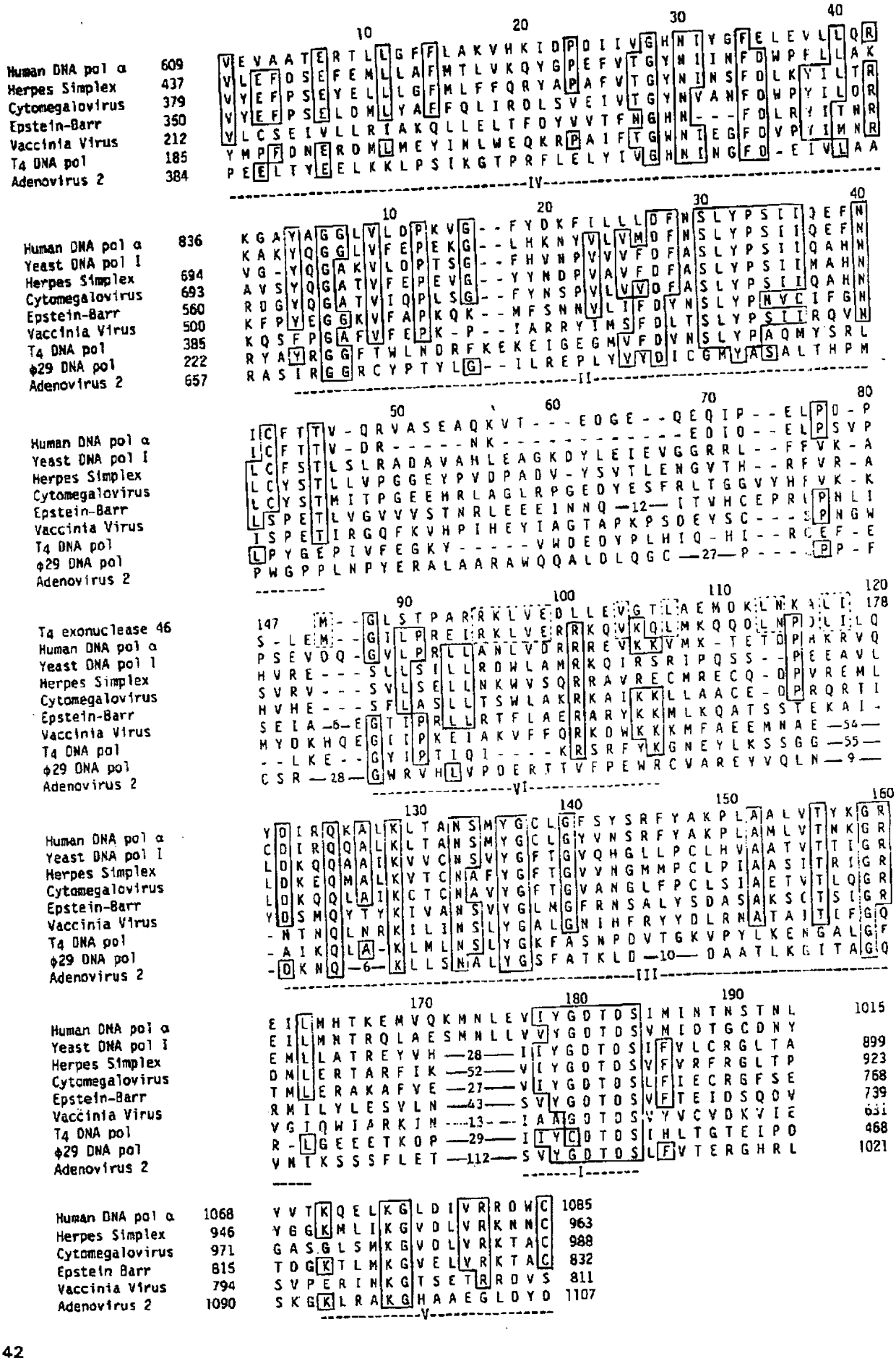
The enzymatic activity and *de novo* protein synthesis of DNA polymerase α both correlate positively with cell proliferation (Bensch *et al.*, 1982; Fry and Loeb, 1986; Thommes *et al.*, 1986). To examine whether the transcriptional expression of polymerase α correlates with *de novo* protein synthesis and the expression of enzymatic activity, a parallel analysis of polymerase α steady-state mRNA levels from quiescent cells stimulated to proliferate was performed. Steady-state mRNA from normal human lung fibroblast cell culture (IMR-90), arrested in quiescent state (G0) by serum deprivation, and from cells activated to proliferate by serum stimulation were compared by Northern blot hybridization (Figure 4B). The steady-state message increases >20-fold, 18 h after serum stimulation, 6 h prior to the peak of DNA synthesis. Meanwhile, the enzymatic activity per cell also increases ~10-fold (Table II). Therefore, the increase of DNA polymerase α steady-state message following the activation of quiescent cells to proliferate is similar to those observed with several other genes involved in DNA synthesis, that undergo transient expression.

A comparative study of message levels in normal growing human tissue and transformed cell lines was also performed. With equal amounts of poly(A)$^+$ mRNA, no detectable polymerase α message was found in RNA isolated from a normal 20-week human placenta, in contrast to a readily detectable 5.8 kb signal from human breast carcinoma cell line (MDA4) (Figure 4C). Genomic Southern blots of normal human cells and transformed cells indicate equal gene dosage of human DNA polymerase α (data not shown), indicating that the abundance of the polymerase α message in transformed cells is not due to the amplification of the polymerase α gene.

Similarities with other DNA polymerases

Comparison of the primary amino acid sequence of human DNA polymerase α deduced from the nucleotide sequences of the cDNA clone to sequences derived from DNA polymerases of herpes, Epstein−Barr, cytomegalo, vaccinia, adeno-2 viruses (Gibb *et al.*, 1985; Earl *et al.*, 1986; Larder *et al.*, 1987; Kouzardies *et al.*, 1987), *E.coli* phage T_4 (National Data Base Bank), *Bacillus* phage φ29 (Yoshikawa and Ito, 1982) and yeast DNA polymerase I (Johnson *et al.*, 1985) reveals several regions of marked similarity. Within a 472-amino-acid region of human DNA polymerase α (amino acids 609−1081) six regions are identified that contain extensive similarities among these DNA polymerases. The regions are designated according to the extent of similarity from I to VI with region I being the most similar (Figure 5A). In addition to these six highly conserved regions, the sequence spanning region VI of human DNA polymerase α, amino acids 908−939, shares ~41% similarity with T_4 gene 46 protein, which is an exonuclease. The significance of these similarities is further underscored by the relative location of these regions within the respec-

A



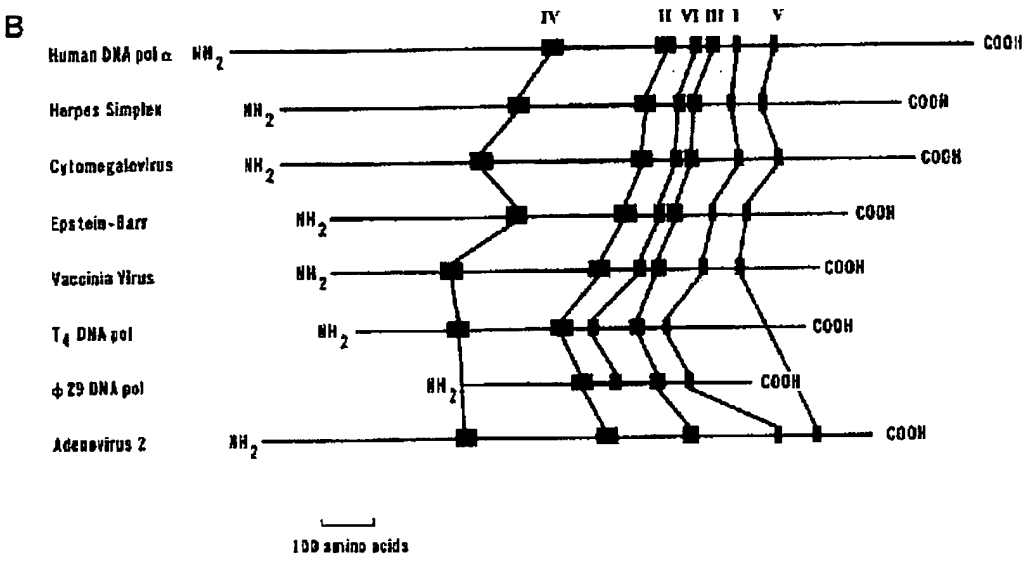


Fig. 5. Conserved regions of human DNA polymerase α and other DNA polymerases. (A) Amino acid sequence similarity between human DNA polymerase α and other DNA polymerases. Amino acid residues 609–1015 from human DNA polymerase α were aligned with amino acid residues derived from other DNA polymerases and T_4 phage exonuclease gene 46 product; identical residues between human polymerase α and other DNA polymerases in five or more sequences are boxed and similar amino acids between polymerase α and T_4 gene 46 protein (exonuclease) are boxed in dashed line. Gaps are indicated by dashes and extensive gaps are marked by the number of amino acids contained within the gap. The designated conserved regions are marked by dashed lines under the amino acid residues. Amino acids 998–1005 of human polymerase α are defined as region I; amino acids 839–878 are region II; amino acids 943–984 are region III; amino acids 609–650 are defined as region IV; region V and VI are amino acids 1075–1081 and 909–926, respectively. (B) Relative spatial arrangemnt of the conserved regions of DNA polymerases. Each DNA polymerase polypeptide is represented by a straight line with NH_2 and COOH denoting the amino and carboxyl termini, respectively. The black bars represent the consensus sequences of each region. Similar regions of each polymerase polypeptide are aligned by vertical lines.

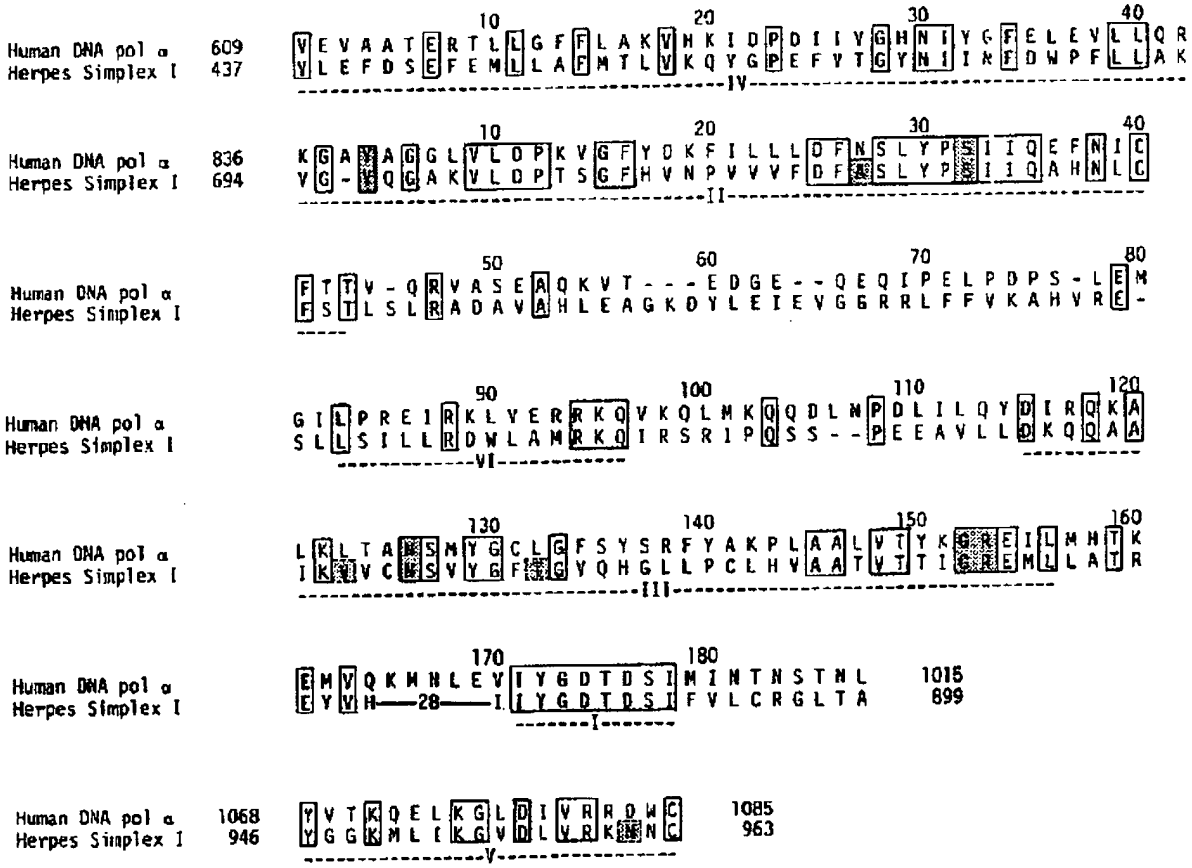


Fig. 6. Conserved sequences between human DNA polymerase α and herpes simplex virus 1 DNA polymerase and predicted deoxynucleotide interacting domains. The six conserved regions between human DNA α and herpes simplex virus 1 were aligned. Identical amino acid residues are boxed. Amino acids that were identified in herpes simplex virus 1 mutants as single amino acid substitution are boxed in shade.

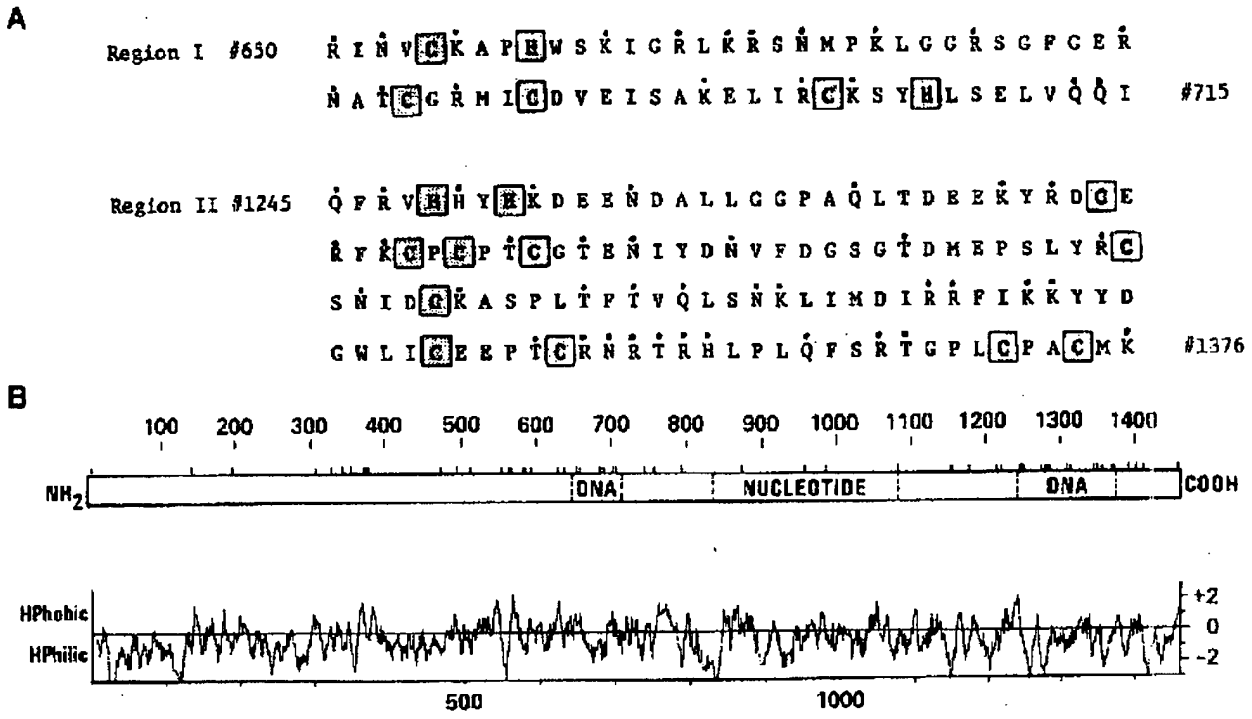


Fig. 7. Amino acid sequences of possible DNA binding regions and schematic summary of putative functional domains of human DNA polymerase α. (A) Possible DNA binding sequences. Two regions of cys/his-rich sequences of human DNA polymerase α are depicted, region I from amino acid 650–715 and region II from amino acid 1245–1376. Cysteine and histidine residues are boxed in shade and amino acids capable of interacting with the phosphate backbone of DNA in the possible loop regions are marked by ●. (B) Schematic representation of the putative functional domains of human DNA polymerase α and hydropathy plot.

tive polypeptides. The six regions are in the same linear spatial arrangement, IV–II–VI–III–I–V, on each polypeptide (Figure 5B). However, the distances between each consensus region in the polypeptides examined are variable.

Sequence comparisons of DNA polymerase β (Zmudzka *et al.*, 1986; Matsukage *et al.*, 1987), terminal transferase (Peterson *et al.*, 1984, 1985) and *E.coli* DNA polymerase I (Joyce *et al.*, 1982) reveal no significant similarity to the conserved regions described above, but a sequence similar to region II is identified in the *dnaE* gene product, the α subunit, of *E.coli* DNA polymerase III (Tomasiewicz and McHenry, 1987).

The presence of these highly conserved domains in replicative DNA polymerases from human to such phylogenetically distant species as bacteriophage T_4 and $\phi 29$ suggest that these DNA polymerases may all be derived from a common primordial gene.

Predicted functional domains

Conservation of these sequences is likely to reflect the need to maintain function. A detailed comparison was made of the amino acid sequences of human DNA polymerase α and herpes DNA polymerase (Figure 6). Extensive sequence similarity in all six regions are identified, with region I having the highest of 87.5%, followed by region II, 60%; region V, 57%; region III, 47%; region IV, 26%; and region VI, 10.5%. Mutations of herpes simplex virus conferring altered anti-viral drug sensitivity have been mapped to several of these conserved regions of the herpes DNA polymerase gene (Knopf *et al.*, 1981; Coen *et al.*, 1983; Quinn and McGeoch, 1985). Most mutants which demonstrate altered sensitivity to the pyrophosphate analog phosphonoacetic acid also exhibit more resistance to the nucleoside analog aphidicolin (Coen *et al.*, 1983; Larder *et al.*, 1987). Sequence analysis of several of these mutants, derived from a single viral strain, confirms that all contain single amino acid substitutions within conserved regions II and III (Knopf, 1987; Larder *et al.*, 1987; Tsurumi *et al.*, 1987) and most recently, another mutant was identified having a single amino acid substitution in region V (J.S.Gibbs, H.C.Chiou and D.M.Coen, personal communication). Mutations conferring altered sensitivity to these drugs are inferred to be at the dNTP and pp_i binding domains. Based on the studies of herpes DNA polymerase mutants, regions II, III and V of human DNA polymerase α could be the essential catalytic domains required for dNTP interaction.

In eukaryotic cells, proteins involved in nucleic acid binding or gene regulation were found to contain cysteine–histidine rich sequences that are potential metal-binding domains which may play an essential role in nucleic acid binding and gene regulation (Miller,J. *et al.* 1985; Berg, 1986). Two regions containing such a motif are found in the DNA polymerase α sequence (Figure 7). One region, amino acids 650–715, contains the sequence Cys-X_3-His-X_{27}-Cys-X_4-Cys-X_{11}-Cys-X_3-His, where X represents amino acids other than cysteine and histidine, and Cys/His-X_n-Cys/His is capable of forming a tetrahedral box structure with an extended protein loop (Figure 7A). The extended loop between residues 659 and 685 contains many amino acids with side chains capable of interacting with the phosphate backbone of DNA (Ohlendrof and Mathews, 1983). Another cys/his-rich sequence within the carboxyl-

terminus of human DNA polymerase α sequence, from amino acid 1245–1376, is identified (Figure 7B). This region contains the sequence His-X_2-His-X_{23}-Cys-X_4-Cys-X-Cys-X_2-Cys-X_{23}-Cys-X_4-Cys-X_{32}-Cys-X_4-Cys-X_{17}-Cys-X_2-Cys, which has the potential to form three tetrahedral box structures defining three extended DNA binding loops as described above. Figure 7C summarizes these possible functional domains.

Discussion

Genetic studies of yeast and several somatic cell lines implies that there are specific regulatory or restriction points in the growth cycle of the cell (Hartwell *et al.*, 1974; Siminovitch and Thompson, 1978). Several genes involved in DNA synthesis such as thymidine kinase (tk) (Groudine and Casimir, 1984; Coppock and Pardee, 1987), thymidylate synthetase (ts) (Storm *et al.*, 1984; Ayusawa *et al.*, 1986) and dihydrofolate reductase (dhfr) (Farnham and Schimke, 1985) undergo transient expression in the cell cycle. DNA polymerase α is the principal enzyme that replicates chromosomal DNA. The steady-state level of polymerase α message increases when cells are activated to proliferate and correlates with the increase of enzymatic activity and *de novo* synthesis of antigenic protein. The concerted increase of these three parameters implies the regulation of the expression of this key DNA replication enzyme is at the transcriptional level. A gene that exhibits transient increase in expression during cell cycle or activation to proliferate could be a regulator of a restriction point, or the target of a cell-cycle or cell-proliferation-specific regulatory signal. Thus far we have only analyzed the proliferation-associated steady-state message of polymerase α. The results imply that expression of polymerase α may be restricted to cells entering the cell cycle. The transient increase in polymerase α mRNA following the stimulation of quiescent cells could be the result of an activational event that renders the cells competent to induce transcription of DNA polymerase α and subsequently initiate DNA synthesis. To further understand the transcriptional regulation of this gene, investigation of the nuclear transcription and steady-state message of DNA polymerase α during activation of quiescent cells to proliferate, as well as within cell cycle, is necessary. The observation of significant amplification of steady-state polymerase α mRNA, *de novo* protein synthesis and enzymatic activity in transformed cells as compared to non-transformed cells, poses the question whether DNA polymerase α is a target for oncogene activation.

Recent studies from various eukaryotic systems indicate that, *in vitro*, the catalytic polypeptide of DNA polymerase α is a polypeptide of 180 kd (Campbell, 1986; Wong *et al.*, 1986). The present amino acid sequence, deduced from the cDNA sequence, demonstrates that the catalytic polypeptide of human DNA polymerase α has a minimum mol. wt of 165 kd. The discrepancy between this value and the 180 kd polypeptide found in DNA polymerase α enzyme preparations suggests either there exists an additional 15 kd of coding sequence further upstream from the putative translation start site, or that a post-translational modification reduces its size. Compared to the yeast DNA polymerase I gene (Johnson *et al.*, 1985) which encodes a biologically functional enzyme of 140 kd, a recombinant human polymerase α of 165 kd may represent the full-length polypeptide. Analysis of a recently isolated genomic clone containing the promoter region of human polymerase α will define the transcriptional and translational initiation sites as well as the primary translation product.

Structure–function studies of DNA polymerase α are prerequisite to understanding the mechanisms by which it interacts with other DNA replication proteins, dNTP and DNA substrates. Analysis of the primary amino acid sequence deduced from the cDNA has identified several structural features which are similar to other nucleic-acid-interacting proteins and DNA polymerases. The identification of cys/his-rich motifs defines putative DNA binding domains. The presence of highly conserved regions sharing similarity with other DNA polymerases and having similar, although not identical, spatial relationships suggests that these regions are needed to maintain essential function. Secondary stucture predictions for each of the three highly conserved regions, I, II and III, indicates that each composes a turn (cleft). Four regions, II, VI, III and I, are localized within a contiguous region that comprises ony 9% of the total length of the polypeptide. This close proximity prompts us to speculate that these domains may form a substrate binding site on the enzyme surface. Experimental evidence based on genetic marker transfer and rescue data of herpes DNA polymerase mutants (Knopf *et al.*, 1981; Coen *et al.*, 1983; Gibb *et al.*, 1985; Quinn and McGeoch, 1985; Tsurumi *et al.*, 1987) implicates regions II, III and possibly region V in substrate dNTP binding and pp_i hydrolysis. Mutations clustered in locations such as region II and III suggest a role for these two sequences in substrate recognition or catalysis. Mutation in region V, which is >100 amino acids apart from regions II and III, results in similar drug resistance phenotypes, suggesting polypeptide folding interactions that form substrate binding sites. The identification of a region II-like sequence in the α subunit of *E.coli* DNA polymerase III further substantiates the functional importance of this region. Thus far there are no regions I, IV and VI mutants isolated from herpes virus. These regions may be the critical domains required for interaction with other accessory proteins in DNA replication or for substrate interaction. The biological function of each of these conserved domains should be definable by site-specific mutagenesis and interchanging host and viral polymerase structural determinants.

It is interesting to note that all DNA polymerases containing these conserved regions identified in this study are replicative enzymes. In addition, except for DNA polymerase α and yeast polymerase I, all of the viral DNA polymerases and the two bacteriophage DNA polymerases have two enzymological activities; a DNA polymerizing activity and a 3′–5′ proof reading exonuclease activity. Since error-free DNA replication is an essential process for the survival of biological organisms, one might expect that the proof reading function would be conserved in this key chromosomal replication enzyme from prokaryotes to eukaryotes. This again raises the issue of the relationship between DNA polymerases α and δ. Does mammalian polymerase α, like *Drosophila melanogaster* polymerase α (Cotterill *et al.*, 1987), have an intrinsic but cryptic 3′–5′ exonuclease activity in the catalytic polypeptide, detectable only when separated from other subunits? Enzymological characterization of the functionally expressed polymerase α catalytic polypeptide should provide an answer to this issue.

The lack of these six conserved sequences in *E.coli* polymerase III, the chromosomal replicative DNA polymerase,

suggests DNA polymerase α and *E.coli* polymerase III evolved from different ancestral genes. These six similar regions in replicative DNA polymerases are conserved to phylogenetically distant species; however, they are notably absent in several eukaryotic DNA polymerizing enzymes: DNA polymerase β (Zmudzka *et al.*, 1986; Matsukage *et al.*, 1987), terminal transferase (Peterson *et al.*, 1985), retroviral reverse transcriptase (Kamer and Argos, 1984), or prokaryotic *E.coli* DNA polymerases I (Joyce *et al.*, 1982) and III (Tomasiewicz and McHenry, 1987). This suggests that there is a class of DNA polymerases which are all replicative DNA polymerases, containing these conserved regions and sharing a primordial archetype.

Materials and methods

General methods

Preparation of plasmid, restriction enzyme digestions and agarose gel electrophoresis were performed as described by Maniatis *et al.* (1982). DNA probes were labeled with ^{32}P by the methods of Feinberg and Vogelstein (1983), synthetic oligonucleotide probes were labeled by T_4 polynucleotide kinase as described by Maniatis *et al.* (1982). Anti-sense ribonucleotide probe was prepared as described (Melton *et al.*, 1984).

Cell lines

KB is a human epidermoid carcinoma cell line; IMR90 is a normal human fetal lung fibroblast cell line; 293 is a human embryonic kidney cell line transformed with sheared adenovirus DNA; MDA4 is transformed human cell line propagated from human mammary carcinoma. These cell lines described above were from ATCC, Rockville, MD. GM1604 is a normal human fetal lung fibroblast cell line and GM1202A is human cell line containing 4X chromosomes, both these cell lines are from NIGMS Human Cell Repository, Camden, NJ. TNHF is a normal human fibroblast primary culture from neonatal foreskin developed by Dr Eric Stanbridge of University California, Irvine, CA. All rodent–human somatic hybrids used in gene mapping were as previously described (Wang *et al.*, 1985).

Isolation of human DNA polymerase α catalytic polypeptides

DNA polymerase α antigen polypeptides from six 18 l cultures of human KB cells (3.5×10^5 cells/ml) were purified with a monoclonal IgG (SJK287)–Sepharose 4B column as described (Wong *et al.*, 1986). The polypeptides were suspended in 0.36 M Tris–HCl, pH 8.6, 3.3 mM EDTA, 8 M urea and then reduced for 3 h at 37°C under N_2 with 10 mM DTT. The reduced polypeptides were alkylated with 22 mM iodoacetic acid at 4°C for 1 h and dialyzed in 50 mM NH_4HCO_3, 0.01% SDS. The dialyzed, reduced and alkylated DNA polymerase α protein was lyophilized, resuspended in 100 mM $NaPO_4$, pH 6.5, and 0.1% SDS and heated at 75°C for 10 min. These polypeptides were then purified by HPLC through two coupled gel permeation columns (TSK 3000, 7.5 × 300 mm) in 100 mM $NaPO_4$, pH 6.5, and 0.1% SDS at a flow rate of 0.5 ml/min. The absorbance of the eluate was monitored at 280 nm. Fractions containing the 180–140 kd DNA polymerase α catalytic polypeptides were dialyzed in 50mM NH_4HCO_3 containing 0.01% SDS and lyophilized.

Peptide sequence analysis

Human DNA polymerase α catalytic polypeptides (500 pmol), isolated as described above, were resuspended in H_2O and ethanol-precipitated twice to remove excess SDS from the samples. The polypeptides were then resuspended in 0.1 M NH_4HCO_3, 10 mM $CaCl_2$ and digested with 2 μg of TPCK treated trypsin at room temperature for 20 h. The trypsin digested peptides were first separated on an Aquapore RP300 (2.1 × 220 mm, Brownlee Lab) HPLC column equilibrated in 0.1% trifluoroacetic acid. A linear gradient from 0–60% acetonitrile was run over 45 min at 0.2 ml/min. Absorbance at 220 nm was monitored by Spectraflow 755 Variable Wavelength detector. Selected peptides peaks were further purified by an RP300 (1 × 100 mm) column equilibrated in 50 mM ammonium acetate, pH 6.5. A linear gradient of 0–75% acetonitrile was run over 30 min at 0.08 ml/min and absorbance monitored at 215 nm. Each of the separated peptides was subjected to automated Edman degradation performed on a model 470A gas phase sequencer with on-line PTH amino acid anlaysis (Model 120A) (Hunkapiller *et al.*, 1983).

Oligodeoxynucleotide probes

Single long anti-sense oligonucleotide probes were designed according to Lathe (1985) and were synthesized on an Applied Biosystems model 380A oligonucleotide synthesizer.

cDNA cloning

Ninety μg of poly(A)$^+$ mRNA from early mid-log human KB cells was heated at 65°C for 1 min and loaded onto a 5.3 ml sucrose gradient of 5–25% containing 100 mM NaCl, 10 mM Tris–HCl, pH 7.4, 1 mM EDTA and 0.1% SDS. Centrifugation was carried out at 52 300 *g* for 25 h at 5°C and fractionated into 20 fractions. mRNA samples of each fraction oligonucleotide probes. Hybridization conditions used were 6 × SSPE, 0.1% SDS 100 μg/ml *E.coli* tRNA. Washing conditions were 2 × SSPE, 0.1% SDS. Temperature of hybridization and washing depended on the individual oligonucleotide probe used. Stringency of hybridization and washing of each individual oligonucleotide probe was based on T_m (melting temperature) and T_w (washing temperature) values estimated at >85% probe–target homology (Lathe, 1985).

Screening of 1×10^5 colonies of this size-selected library yielded a single distinct positive clone designated as pcD-KB*polα*, which hybridizes with oligodeoxynucleotides T264, T265 and T25 (Table I). Sequence analysis of pcD-KB*polα* (Figure 2) indicates that it contains a 2893-bp cDNA insert with an open reading frame of 1865 bp terminated by a stop codon and followed by a 1028-bp non-coding region. In this 1865-bp coding sequence there are four regions of deduced amino acid sequences that are perfectly homologous to the previously determined amino acid sequences, T264, T265, T25 and T9 (Table I). The 3′-non-translated region contains several in-frame stop codons, and the consensus polyadenylation signal AATAAA (Proudfoot and Brownlee, 1976) 13 nucleotides upstream from the polyadenylation tail. This indicates that pcD-KB*polα* contains the 3′-end of the cDNA for human DNA polymerase α. To extend this truncated cDNA clone the 5′-most restriction fragment of pcD-KB*polα*, *PstI/Hin*dIII, was used to screen 2×10^6 phage of a human pre-B cell cDNA library (E1 library) constructed in λgt10 (Cleary *et al.*, 1986). The very 5′-terminal restriction fragments of the newly extended cDNA clones were used to further screen the E1 library. The complete set of overlapping clones was sequenced in both directions as described (Dale *et al.*, 1985) and reassembled.

Genomic DNA and RNA blot hybridization

Genomic DNA hybridization. Five or 10 μg of human genomic DNA were digested with *Eco*RI. DNA blot hybridization was carried out with 50 ηg of ^{32}P-labeled *PstI/PstI* 700-bp fragment of pcD-KB*polα* (10^9 c.p.m./μg). Hybridization was at 6 × SSC, 50 mM $NaPO_4$, pH 7.0, 5 × Denhardt solution, 100 μg boiled and sonicated salmon sperm DNA, 50% formamide and 10% dextran sulfate at 42°C. The blot was washed in 0.2 × SSC, 0.1% SDS at 65°C.

RNA hybridization. Polyadenylated mRNA was analyzed on a 1% agarose gel in 6.3% formaldehyde. Northern blot hybridization was carried out with 50 ηg of ^{32}P-labeled restriction fragment of cDNA (10^9 c.p.m./μg). Hybridization and wash were as described above.

Amino acid sequence of other DNA polymerases

Nucleotide sequence from yeast DNA polymerase I was determined from sequence analysis of a *Eco*RI/*Hin*dIII restriction fragment of yeast DNA polymerase I gene (Johnson *et al.*, 1985). Other viral DNA polymerase sequences and T_4 gene 46 sequence were derived either from published data or National Data Base Bank of the National Biomedical Research Foundation.

Acknowledgements

We thank Dr T.Yokota and Mrs K.Yokota for advice and construction of size fractionated pcD KB cDNA library; Dr M.L.Cleary for the human pre-B cell cDNA library and cDNA cloning advice; A.Johnson for tissue culture assistance; Drs M.Synder and R.Davis for the yeast DNA polymerase I gene; Drs E.C.Friedberg and D.A.Clayton for helpful comments on the manuscript. This research was supported by a grant from the US National Cancer Institute (CA 14835) to T.S.F.W. and by a gift from the Donald E. and Delia B. Baxter Fund. S.W.W. is a predoctoral fellow and A.F.W. is a postdoctoral fellow supported by NIH training grants CA09302 and CA09151, respectively.

References

Ayusawa,D., Shimizu,K., Koyama,H., Kaneda,S., Takeishi,K. and Seno,T. (1986) *J. Mol. Biol.*, **190**, 559–567.

Bensch,K.G., Tanaka,S., Hu,S.-Z., Wang,T.S.-F. and Korn,D. (1982) *J. Biol. Chem.*, **257**, 8391–8396.

Berg,J.M. (1986) *Science*, **232**, 485–487.

Bravo,R., Frank,R., Blundell,P.A. and Macdonald-Bravo,H. (1987) *Nature*, **326**, 515–517.

Byrnes,J.J., Downey,K.M., Black,V.L. and So,A.G. (1976) *Biochemistry*, **15**, 2817–2823.

Campbell,J.L. (1986) *Annu. Rev. Biochem.*, **55**, 733–771.

Cleary,M.L., Smith,S.D. and Sklar,J. (1986) *Cell*, **47**, 19–28.

Coen,D.M., Furman,P.A., Aschman,D. and Schaffer,P.A. (1983) *Nucleic Acids Res.*, **11**, 5287–5297.

Coppock,D. and Pardee,A.B. (1987) *Mol. Cell. Biol.*, **7**, 2925–2932.

Cotterill,S.M., Reyland,M.E., Loeb,L.A. and Lehman,I.R. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5635–5639.

Dale,R.M.K., McClure,B.A. and Houchins,J.P. (1985) *Plasmid*, **13**, 31–40.

Earl,P.L., Jones,E.P. and Moss,B. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3659–3663.

Farnham,P.J. and Schimke,R.T. (1985) *J. Biol. Chem.*, **260**, 7675–7670.

Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–13.

Fry,M. and Loeb,L.A. (1986) in *Animal Cell DNA Polymerases*. CRC Press, Boca Raton, FL. pp. 13–60.

Gibb,J.S., Chiou,H.C., Hall,J.D., Mount,D.W., Retondo,M.J., Weller,S.K. and Coen,D.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 7969–7973.

Groudine,M. and Casimir,C. (1984) *Nucleic Acids Res.*, **12**, 1427–1446.

Hartwell,L.H., Culotti,J., Pringle,J. and Reid,B. (1974) *Science*, **183**, 46–51.

Hunkapiller,M.W., Hewick,R.M., Dreyer,W.J. and Hood,L.E. (1983) *Methods Enzymol.*, **91**, 399–413.

Ikegami,S., Toguchi,T., Ohasi,M., Oguro,M., Nagano,H. and Mano,Y. (1978) *Nature*, **275**, 458–459.

Johnson,L.M., Snyder,M., Chang,L.M.S., Davis,R.W. and Campbell,J.L. (1985) *Cell*, **43**, 369–377.

Joyce,C.M., Kelley,W.S. and Grindley,N.D.F. (1982) *J. Biol. Chem.*, **257**, 1958–1964.

Kamer,G. and Argos,P.K. (1984) *Nucleic Acids Res.*, **12**, 7269–7283.

Knopf,K.W., Kaufman,E.R. and Crumpacker,C. (1981) *J. Virol.*, **39**, 746–757.

Knopf,K.W. (1987) *J. Gen. Virol.*, **68**, 1429–1433.

Kornberg,A. (1980) *DNA Replication*. W.H.Freeman, San Francisco, CA.

Kornberg,A. (1982) *DNA Replication, 1982 Supplement*. W.H.Freeman, San Francisco, CA.

Kouzardies,T., Banker,A.T., Satchwell,S.C., Weston,K., Tomlinson,P. and Barrell,B.C. (1987) *J. Virol.*, **61**, 125–133.

Kozak,M. (1981) *Nucleic Acids Res.*, **9**, 5233–5252.

Larder,B.A., Kemp,S.D. and Darby,G. (1987) *EMBO J.*, **6**, 169–175.

Lathe,R. (1985) *J. Mol. Biol.*, **183**, 1–12.

Lee,M.Y.W.T., Tan,C., Downey,K.M. and So,A.G. (1984) *Biochemistry*, **23**, 1906–1913.

Li,J.J. and Kelly,T.J. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 6973–6977.

Li,J.J. and Kelly,T. (1985) *J. Mol. Cell. Biol.*, **5**, 1238–1246.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York.

Matsukage,A., Nishikawa,K., Ooi,T., Seto,Y. and Yamaguchi,M. (1987) *J. Biol. Chem.*, **262**, 8960–8962.

Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T., Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.

Miller,J., McLachlan,A.D. and Klug,A. (1985) *EMBO J.*, **4**, 1609–1614.

Miller,M.R., Ulrich,R.G., Wang,T.S.-F. and Korn,D. (1985) *J. Biol. Chem.*, **260**, 134–138.

Miller,M., Seighman,C. and Ullrich,R.G. (1986) *Biochemistry*, **24**, 7440–7445.

Murakami,Y., Yasuda,H., Miyazawa,H., Hanaoka,F. and Yamada,M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1761–1765.

Murakami,Y., Wobbe,C.R., Weissbach,L., Dean,F.B. and Hurwitz,J. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 2869–2873.

Ohlendorf,D.H. and Mathews,B.W.A. (1983) *Rev. Biophys. Bioeng.*, **12**, 259–284.

Okayama,H. and Berg,P. (1982) *Mol. Cell. Biol.*, **2**, 161–170.

Okayama,H. and Berg,P. (1983) *Mol. Cell. Biol.*, **3**, 280–289.

Perlich,G., Kostura,M., Marshak,D.R., Mathews,M.B. and Stillman,B. (1987a) *Nature*, **326**, 471–475.

Perlich,G., Tan,C., Kostura,M., Mathews,M.B., So,A.G., Downey,K.M. and Stillman,B. (1987b) *Nature*, **326**, 517–520.

Peterson,R.C., Chueng,L.C., Mattaliano,R.J., White,S.J., Chang,L.S.M. and Bollum,F.J. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4363–4367.

Peterson,R.C., Cheung,L.C., Mattaliano,R.J., White,S.T., Chang,L.M.S. and Bollum,F.J. (1985) *J. Biol. Chem.*, **260**, 10495–10502.

Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211–214.

Quinn,J.P. and McGeoch,D.J. (1985) *Nucleic Acids Res.*, **13**, 8143–8163.

Siminovitch,L. and Thompson,L.H. (1978) *J. Cell Physiol.*, **95**, 361–366.

Stillman,B.W. and Gluzman,Y. (1985) *Mol. Cell. Biol.*, **5**, 2051–2060.

Storm,R.K., Ord,R.W., Greenwood,M.T., Mirdamadi,B., Chu,F.K. and Belfort,M. (1984) *Mol. Cell. Biol.*, **4**, 2858–2864.

Thommes,P., Reiter,T. and Knippers,R. (1986) *Biochemistry*, **25**, 1308–1314.

Tomasiewicz,H.G. and McHenry,C.S. (1987) *J. Bact.*, in press.

Tseng,B.Y. and Ahlem,C.N. (1983) *J. Biol. Chem.*, **258**, 9845–9849.

Tsurumi,T., Maeno,K. and Nishiyama,Y. (1987) *J. Virol.*, **61**, 388–394.

Wang,T.S.-F., Hu,S.-Z. and Korn,D. (1984) *J. Biol. Chem.*, **259**, 1854–1865.

Wang,T.S.-F., Pearson,B.E., Soumalainen,H.A., Mohandas,T., Shapiro,L.J., Schroder,J. and Korn,D. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 5270–5274.

Wobbe,C.R., Dean,F., Weissbach,L. and Hurwitz,J. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 5710–5714.

Wong,S.W., Paborsky,L.R., Fisher,P.A., Wang,T.S.-F. and Korn,D. (1986) *J. Biol. Chem.*, **261**, 7958–7968.

Yoshikawa,H. and Ito,J. (1982) *Gene*, **17**, 323–325.

Zmudzka,B.Z., SenGupta,D., Matsukage,A., Cobianti,F., Kumar,P. and Wilson,S.H. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5106–5110.

Received on October 12, 1987